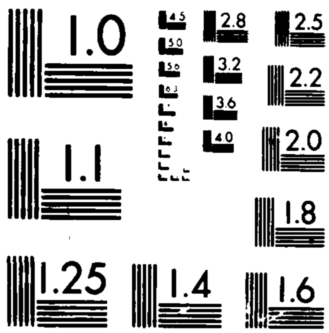
AD-A193 392 COGNITIVE AND NEURAL BASES OF SKILLED PERFORMANCE(U) 1/2
NEW YORK UNIV N Y L KAUFMAN 04 OCT 87 AFOSR-TR-87-1710
F49620-86-C-0131

UNCLASSIFIED F/G 5/8 NL


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0

~~FILE COPY~~

AD-A193 392

REPORT DOCUMENTATION PAGE

DTIC FILE COPY

DTIC ELECTE APR 0 4 1988 S D

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR- 87-1710
6a. NAME OF PERFORMING ORGANIZATION New York University	6b. OFFICE SYMBOL (If applicable)
7a. NAME OF MONITORING ORGANIZATION Air Force Office of Scientific Research	
6c. ADDRESS (City, State and ZIP Code) Departments of Physics and Psychology, 4-6 Washington Place, New York, NY 10003	7b. ADDRESS (City, State and ZIP Code) Building 410, Bolling AFB, DC 20332-6448
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR	8b. OFFICE SYMBOL (If applicable) NL
9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER F49620-86-C-0131	
8c. ADDRESS (City, State and ZIP Code) Building 410, Bolling AFB, DC 20332	

10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
61103D	3484	A4	

11. TITLE (Include Security Classification) Cognitive and Neural Bases of SkilledPerformance

12. PERSONAL AUTHOR(S) Lloyd Kaufman

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
ANNUAL TECHNICAL	FROM 9/86 TO 9/87	October 4 1987	19

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	Cognition, perception skilled performance, neuromagnetism

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Research in seven interrelated areas has been carried out during this first year of support. A major advance in the capability of neuromagnetic monitoring of higher levels of brain function was achieved with the installation of a large magnetically shielded room for the sensing system. This together with a computer-controlled adaptive filter dramatically improved sensitivity to slowly-varying brain signals. A new analysis technique based on a multidimensional signal space has also been developed to characterize the neural configurations that give rise to differing field patterns, without restricting consideration to only the simplest that can be attributed to an equivalent current dipole. Successful application of these concepts to studies of the alpha rhythm also demonstrated a major advantage in using elements of the covariance matrix across sensors to reduce interference from extraneous neural sources, thereby improving the signal-to-noise ratio by an order of magnitude. In preparing for a series of collaborative studies, various mehtods for producing visual displays in the shielded room were evaluated, and one based on a single-lens TV projection system was chosen. The system can be controlled by a computer peripheral to the HP9000/550 used to record and analyze neuromagnetic data. Software was successfully

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE NUMBER (Include Area Code)	22c. OFFICE SYMBOL
Dr. Alfred Fregly	202-767-5021	NL

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

28 OCT 1987

ABSTRACT CONTINUED

developed so that either an IBM/PC-compatible or Amega computer can be used for this purpose. An Adage 3000 frame buffer was obtained for psychophysical studies, and programs were developed to present a display where only color or only luminance changes are produced. A coordinated study of physiological and psychological response to changing direction of visual pattern motion began with an upgrade of computer memory in the visual neuroscience laboratory and development of dynamic stochastic random motion displays, based on a pattern of random dots. These have the virtue of testing sensitivity to motion while leaving factors such as spatio-temporal contrast energy constant. Physiological studies of the extrastriate visual area of the medial temporal lobe revealed cells with high sensitivity to this display. Perceptual studies with human subjects of the spatial and temporal determinants of sensitivity to motion are nearly complete. In initial work to combine neuromagnetism with psychological methods for the study of memory and related cognitive proceses, a paradigm has been successfully developed to present verbal stimuli on a video screen using a selected set of words that are either highly imageable or highly abstract. A simple task was set up to demonstrate the difference between these words in memory. A new technique was also developed to enhance the memorability of the items. Another research program on the dimensional structure of knowledge states has as a goal the development of computerized procedures to assess the state of knowledge of an individual in a particular domain of information. Initial investigations of a biorder dimension showed the importance of errors that are inherent in data; consequently, the extension of this work to probabilistic situations is underway, including consideration of the advantages of random utility models. Research on visual learning and measurement has addressed the question as to what kind of training maximizes perceptual learning and what kind episodic learning. A visual learning protocol based on a picture fragment completion task has been developed and tested. Four experiments were carried out in which training was manipulated and the effects of on perceptual learning and episodic learning was measured. Control studies carried out early during the testing show that there is high variability in the level of difficulty from item to item. This problem is being overcome by use of a normative experiment conceived for this purpose.

AFOSR-TR- 87-1710

ANNUAL TECHNICAL REPORT
Cognitive and Neural Bases of Skilled Performance
A UNIVERSITY RESEARCH INITIATIVE

Lloyd Kaufman, Principal Investigator
New York University

1. Introduction

Pursuant to the requirements of the Contract No. F49620-86-C-0131, between AFOSR and New York University we submit this report which describes the work accomplished during the first year of our URI project.

The present project began in September 1986 and had four basic components. These included the improvement of university facilities to permit the conduct of advanced research that would be in the interest of the Department of Defense, the actual conduct of such research, the training of students and postdoctoral personnel to conduct the research, and the coordination of research and training efforts with Department of Defense laboratories which would be to the mutual denefit of the University and the DoD. Hence, this report describes our progress to date in all of these areas.

Before entering into a detailed description of our progress, it is important that we describe the nature of the particular URI effort at New York University. The title of our URI program suggests that its focus would be the study of perception, cognition and neural factors relevant to skilled performances. The centerpiece of our proposal was our already highly advanced capabilities in the non-invasive techniques of neuromagnetism, which permits the study of brain activity relevant to perception, cognition and motor behavior. In our original proposal we focussed on the fact that it is now time to proceed towards a greater integration of the work in neuromagnetism with work in cognition and performance. We shall assume that the term "cognition" includes such classic areas of study as memory, and the perception of form, space, speech, motion, and color. It is also presumed to include attention, decision making, and to be basic to the execution of skilled performances of many kinds. The faculty at NYU is already making outstanding contributions in research as well as training of scientists to work in these areas. The URI gave us an opportunity to make a major step toward integration of work in these diverse areas, especially with regard to integrating classic approaches to cognition with the emerging methods for non-invasively pinpointing relevant neural mechanisms in the human brain. However, we did not propose to curtail any of the other work being conducted on an independent basis by the various co-principal investigators. Rather, we sought to complement these efforts, and to lead the investigators into areas of fruitful collaboration. Moreover, we believed that the URI would make it possible for us to enhance the training of students in multidisciplinary areas which had previously not been available to them. Finally, we felt that cooperation with DoD laboratories would be to our mutual advantage. With all of this in mind, we may now turn to a report on our progress toward achieving these objectives.

88 4 4 023

87 1[redacted]

2. Activities in the Neuromagnetism Laboratory

2.1. Magnetically Shielded Room

Performance of our 5-sensor neuromagnetometer system (Williamson et al. 1984) was dramatically improved with the installation of a magnetically shielded room (MSR), whose purchase was made possible with the URI grant. Our experience with unshielded measurements showed that below about 3 Hz, environmental noise rises above the instrinsic sensor noise and shows an approximately inverse power-law dependence on frequency. This is a spectral region of particular interest for studies of higher levels of brain function, such as the P300 signal (Okada et al. 1982). Magnetic shielding offers a solution to this problem, as first demonstrated by Cohen et al. (1970) when recording magnetocardiograms in a multi-layer chamber at the Francis Bitter National Magnet Laboratory at M.I.T. While that shield was constructed with a shape approximating that of a sphere, recently constructed shields (Mager 1981; Keihä et al. 1982) show that adequate performance can be obtained with a cubic or rectangular shape and four or five shells of magnetic shielding. Such a geometry is particularly advantageous because it provides greater usable interior space for a given exterior size. Eddy-current shielding has also been used to some advantage with rooms fabricated from thick aluminum plate (Zimmerman, 1977), but the effectiveness of this shielding is markedly reduced below a few hertz frequency, where the environmental noise is often most severe.

2.2. Installation of a New Shielded Room

During the summer of 1987 a two-shell MSR was installed in our laboratory at the Washington Square Center of New York University. It was erected in room 925, on the 9th floor of Meyer Hall of Physics. This MSR was prefabricated by Vacuumschmeltze GmbH of West Germany, and the parts were assembled on site under the supervision of the company's personnel. The interior has floor dimensions of 3 m × 4 m and a height of 2.4 m. As such it is comparable to, or exceeds in size, conventional rf-shielded rooms and Faraday shields commonly used in EEG research. Such a generous space is important for subject comfort and safety, and to make it convenient for the experimenter to assemble visual displays.

Our MSR consists of an inner shield of mu-metal™ mounted on 8-mm thick aluminum plate that serves as an eddy-current magnetic and radio-frequency shield. It has a conventional stretch-wall covering over interior and exterior for protection and attractive appearance. The inner wall montage is supported on the outside by a stiff aluminum framework of 15-cm thickness, and the outer surface of the framework is covered by a second mu-metal shield and stretch-wall partitions. A single door, balanced on heavy-duty ball bearings, provides magnetic continuity for the two layers of magnetic shielding. The door is secured by a wheel-type of locking mechanism that can be opened from either the inside or outside.

The strength of the MSR frame is sufficient that when a 300 kg weight is attached to the center of the ceiling, the sag is less than 0.3 mm. This makes is possible for the ceiling to support a pair of aluminum rails used to suspend a gantry for supporting the 5-sensor system. Furthermore, the stiff framework will support the gantries for the pair of CryoSQUID sensors (developed with support from a DOD-University Research Instrumentation Grant from the Air Force Office of Scientific Research). The vertical support for each gantry will be bolted to the corners farthest from the door. With this arrangement of overhead gantry and two side-reaching gantries, the three dewars (holding a total of 7 sensors) can be positioned independently over three different areas of the subject's head.

Air circulation for the inside of the MSR is supplied by a duct fed directly by the laboratory building supply. Helium gas slowly vented by the boiling liquid helium in the 5-sensor dewar is conducted by tygon tubing to the outside to avoid any possible build-up. Vents high on the front wall permit free-flow of air in any event. Access ports are provided for air circulation, a projection video display, and filtered electrical leads.

2.3. MSR Performance

The attenuation provided by this MSR was evaluated by R.T. Johnson and J.R. Marsden by placing an electromagnet about 5 m or more from the wall of the room with its axis vertical. A SQUID magnetometer positioned at the center of the room detected the interior field strength when fields of various frequencies were applied. After a 60-Hz shaking field was applied, the steady earth's field is attenuated by a factor of 10^3. However, shielding is less effective for very low-frequency fields as is usual with such materials, being only about 30 db. The effect of eddy-current shielding becomes apparent above about 10^{-1} Hz, and the attenuation rises to a value of about 10^4 at 10^2 Hz with a tendancy toward saturation at higher frequencies. The shielding is not quite as effective as for MSRs having three[17] or more[15] separated layers of magnetic shielding, but the present room is considerably less expensive and has greater interior space. Furthermore, when second-order gradiometers are used as detection coils in the room, the noise spectrum is similar to that of the more expensive MSRs.

The improvement in noise level provided by the MSR is indicated by data obtained for the noise spectrum. The field spectral density was measured by J. Shang and B. Schwartz of our laboratory with the center detection coil of our 5-sensor system placed near the center of the room and oriented vertically. This sensor, has a second-order gradiometer as its detection coil, with a baseline between adjacent coils of 4 cm. In the absence of the MSR the low-frequency ambient noise first became apparent over the intrinsic sensor noise at a frequency of 25 Hz (Williamson et al. 1984), increasing steadily with decreasing frequency. With the MSR it becomes apparent at the considerably lower frequency of about 0.8 Hz (Buchanan et al. 1987). This spectrum may be compared with the performance of our 5-sensor system when reference signals from three other SQUID sensors in the dewar are appropriately amplified and subtracted from each of the signals. These three have magnetometer detection coils that are oriented to monitor three orthogonal components of the ambient field. With such "electronic balancing" of the signal we were able in our unshielded environment to reduce the onset of the ambient noise in the spectrum to about 3 Hz (Williamson et al. 1984). Thus, the MSR performs better than the method of electronic balancing that we had previously relied on.

2.4. Electronic Noise Cancellation

The availability of the reference sensors in the dewar made it possible for us to carry noise reduction one step farther. Simple fixed electronic balancing techniques have already been applied with success for unshielded measurements . Dr. S.E. Robinson of BTi developed a new computer-based approach that in many practical applications markedly reduces excess low frequency noise (Williamson et al. 1988). The technique uses the same computer-controlled attenuators to adjust the reference levels previously employed from electronic balancing, but in this case they are adjusted by a computer algorithm so as to remove the *correlated* portion of the noise from each signal channel. This reduces the noise spectral density by a factor of 5 at 0.1 Hz at the Neuromagnetism Laboratory. With such electronic noise cancellation, the noise level is sufficiently low to permit operation of the SQUID sensors in a dc-coupled mode to monitor very low-frequency activity. It represents state-of-the-art performance.

2.5. Signal Space Characterization of Neural Sources

In collaboration with Dr. R.J. Ilmoniemi and Dr. W.E. Hostetler, we have developed new methods to analyze multichannel neuromagnetic recordings of spontaneous brain activity that avoid specific assumptions concerning the nature of the sources (Ilmoniemi et al. 1988). This approach can be applied to studies of a variety of brain signals, such as sensory related activity, alpha rhythm, and interictal epileptic activity, for the purposes of classification and analysis, even if the field pattern over the scalp may not be characterized as that of a single current dipole. We developed this approach by applying these methods to the study of the alpha rhythm in human subjects. This was an attractive strategy because the signal is strong and it could be measured conveniently with the 14-sensor system at the NYU Medical Center. While not a part of this present work, we have the intention of extending this research to determine the actual positions for the sources of alpha spindles. Since a current dipole determined magnetically is characterized by 5 parameters, our 5-sensor system would not be appropriate since its data would not overdetermine the situation.

2.6. Background on the Alpha Rhythm

By definition (Chatrian et al. 1974), the alpha rhythm is brain activity that gives rise to electrical oscillations between 8 and 13 Hz on the occipital scalp and is attenuated by visual stimuli. The cortical origin of alpha activity has been evidenced by studies of potentials at various depths within the cortex of animals (Calvet et al. 1964; Creutzfeldt and Houchin 1974; Frost 1968) and especially by a clear polarity reversal between superficial and depth electrodes in dog (Lopes da Silva and van Leeuwen 1978). These last studies also provided evidence that alpha activity originates in different epicenters from which activity spreads in several directions, rather than originating in a single source and sweeping over a large area of cortex.

Previous neuromagnetic studies (Carelli et al. 1983; Vvedensky et al. 1986) suggest that many sources responsible for magnetically monitored alpha activity are located near or in the visual cortex and that there are time series of the rhythm during periods of strong activity in which the oscillation period is constant. These time series are called *spindles*. The magnetic field pattern during a spindle appears to remain relatively stable. Based on these findings, we adopt the working hypothesis that there are *specific configurations of neurons* that exhibit such oscillatory excitations. We call such oscillations *alphons*. These hypothetical oscillations would be basic units of activity, whose signals add up to form the observed alpha rhythm. Our analysis was developed to test this alphon hypothesis by determining whether the underlying sources of spindles can be said to differ significantly. In particular it would be interesting to determine whether an alphon can be associated with many spindles, indicating that the alphon exhibited repeated oscillations.

A similar analysis can be applied to multi-sensor studies of sensory evoked magnetic fields, where the source of an observed field pattern is characterized by the relative amplitudes of signals in the sensors.

2.7. Signal Space

It is possible to characterize the source configuration of an alphon without the use of any model, such as a current dipole model. We study the kinematics of the alpha rhythm in an n-dimensional "signal" space, as defined by the output amplitudes of the n-sensor system. The lead fields of n detection coils are not generally orthogonal in current space (Hämäläinen and Ilmoniemi 1984) nor do they form a complete set of basis vectors (they span a subspace of the current space). Nevertheless, even without orthogonalizing the set of n basis vectors we may investigate the distribution of activity within this space. (Carrying out an orthogonalization would simply provide a more economical description of the phenomenon.)

A given source configuration corresponds to a specific direction for its *spindle vector* in signal space, as defined by the 14 sensor outputs, and a greater distance from the origin in a fixed direction represents greater source strength. Thus we may ignore source strength *per se* in our discussion if we confine consideration to direction alone. For instance, we can determine whether more than one spindle can be attributed to the same alphon by establishing whether the spindle vector in this space points in the same direction as the spindle from that alphon (the *alphon vector*). Furthermore, it would be possible to determine whether different directions of signal space have characteristic features, such as different alpha frequencies.

2.8. Magnetic Measurements and Interpretation

Alpha activity for two subjects was recorded with the GEMINI system at the Center for Neuromagnetism of the New York University Medical Center. This system has two probes in separate dewars, each with seven second-order SQUID gradiometers. When maintained in fixed positions the system defines a 14-dimensional signal space. The geometry of the detection coils within each probe is identical to that of a 5-sensor probe described by Williamson et al. (1984) except that each central detection coil is surrounded by 6 rather than 4 outrider coils, which lie on a circle 4 cm in diameter. The subject was prone on a bed in a room with subdued lighting, with an unobstructed downward view onto a patterned surface. The probes were positioned over the occipital area of the scalp about 5 cm above the inion and 6 cm to either side of the midline, where the alpha signals were found to be strongest. Data were recorded continuously for 12-second epochs, with the subject's eyes closed; and at fixed intervals recordings were done with eyes open to verify the existence of alpha blocking in this condition and to obtain an estimate of the background noise. The recording bandwidth was 1-50 Hz; for analysis, the passband was narrowed to the alpha band, i.e., 8-13 Hz.

The initial analysis of the data consists of using a computer to automatically detect intervals of strong spindles and picking those that have a stable period and no phase shifts across sensors. Typically, 1 to 5 such spindles were detected during a 12-second epoch. If the alphon hypothesis were valid, these spindles could be due to individual alphons, and the field pattern should remain stable in signal space during each spindle's lifetime. To improve the signal-to-noise ratio in determining spindle amplitudes, the covariance between the signal of a given sensor and each of the other signals was computed (described later in this report in more detail). These covariances were added together and divided by the sum of the other signal amplitudes to obtain the portion of the given sensor's signal that is coherent across sensors. The 14 amplitudes thus obtained define the components of the spindle vector in signal space.

To determine whether the source configurations of two spindles differ significantly, we need only determine whether the angle between their spindle vectors differs significantly from the noise. Conceptually, it may be useful to consider noise for each measurement as represented by an ellipsoid in signal space. It is an ellipsoid rather than a sphere, since generally the instrumental, environmental, and subject noise differ across sensors, and the latter two may also be correlated across sensors. If we imagine such a noise ellipsoid centered on the head of each spindle vector, then when the ellipsoids of two spindle vectors do not have angular overlap, the measurement has indicated the existence of separate source configurations.

All pairs (i,j) of spindle vectors were analyzed to determine whether their angle of separation α_{ij} significantly exceeds the noise. We determined this by projecting in turn the background noise for each of the 14 sensors onto the plane defined by the pair of spindle vectors, and adding these noise vectors to the ith spindle vector with signs for their components that most quickly bring the resultant vector toward the jth spindle vector. The corresponding angle between the ith spindle vector and the resultant when noise is added is called the *noise angle* η_{ij}. In a similar fashion the noise angle η_{ji} for the second spindle vector is computed. The total noise angle is then taken as $\varepsilon_{ij} = [\eta_{ij}^2 + \eta_{ji}^2]^{1/2}$. We introduce the *discrimination ratio* $D_{ij} = \alpha_{ij}/\varepsilon_{ij}$ to represent the angular separation of two spindle vectors compared to the noise. D_{ij} is a measure of the significance of the difference between a pair of spindle sources. All spindles that cannot be distinguished from a given one are called its *neighbors*.

The data for 4 trials with subject RJI and 3 trials for subject FAL were analyzed as described above. Each trial contained 18 twelve-second epochs, the epochs being separated by 8 seconds. In each trial, 26 to 35 spindles were found, based on the criteria that a spindle has to exceed a minimum amplitude of 0.5 pT and a minimum duration of 500 ms, and that the oscillations in different sensors must be coherent. In addition, phase stability within each channel during a spindle was required. Typical spindles had durations under 2 sec and peak-to-peak amplitudes of about 1 pT. The range of observed frequencies for RJI was from 10.0 to 11.2 Hz and for FAL from 9.6 to 10.8 Hz.

Typical angles between spindle vectors in signal space were 10-30 deg. With our signal-to-noise ratio of about 8, we found for both subjects that almost every spindle could be distinguished from each of the others in a trial with a discrimination ratio $D_{ij} = 2.0$. Therefore, each spindle characterizes a different alphon at this discrimination level. In other words, the spindle vectors were sufficiently different in orientation that the angle between any two vectors was more than 2 times the total noise angle for the two. The same was true with the more stringent requirement $D_{ij} = 4.0$ for FAL, but at this level for RJI only 9-12 alphons were needed to explain each set of 27 to 35 spindles in a trial. In other words, if we require vectors to be separated by more than 4 times the total noise angle for them to be considered distinct, fewer vectors were significantly different from each other. The observed spindles could be explained by only one-third that number of alphons. A typical alphon could account for 3 to 10 spindles (called the *cohorts* of the alphon), and each spindle could be explained by any of 1 to 6 alphons (called the *candidates* for the spindle). Successive spindles were often less distinguishable from each other than spindles separated by longer times. Generally, as the signal-to-noise ratio increases so that spindles may be better discriminated from each other, we expect to have fewer cohorts for each alphon and fewer candidates for each spindle.

Spindles that are cohorts of a given alphon for RJI were often found to have differing frequencies. Therefore, either an alphon is not limited to oscillating at a given frequency each time it is excited, or with improved signal-to-noise ratio we could be able to distinguish between these spindles of differing frequency.

This study shows it is possible to distinguish between most of the sources of observed spindles using the spindle vector representation. For the present subjects, some spindles that are cohorts of an alphon are found to have differing physical properties such as frequency. This implies that to within the resolution of our measurements the underlying neural excitation with fixed geometry can be modulated. Our results indicate that the alpha rhythm is generated by a large number, or possibly a continuum, of different source configurations. It remains to be seen whether most or even any of these are locally oscillating portions of cortex so that they might be modeled by current dipoles.

Analyses of this type based on a signal space concept can be applied to classify the sources of signals from other types of neural excitations, such as interictal epileptic activity and sensory evoked responses. One advantage is that this method is not computationally demanding, and model-specific analyses such as high-precision source localization with realistic numerical models of the subject's head need be performed only once for each class, such as for an alphon representing *all* of its cohorts.

2.9. The Effect of Correlation on Noise

The preceding study was possible because we applied a correlation technique to improve the signal-to-noise ratio for the recorded spindles. Dr. R.J. Ilmoniemi developed a program to compute the covariance matrix between the 14 recorded time series representing a given alphon. The element representing the covariance between channel i and j is defined as:

$$C_{ij} = T^{1} \Sigma_t X_i(t) X_j(t),$$

where the summation over N sampling instants runs from $t = \tau$ to T, with $T = N\tau$. The recording in channel i can be expressed as:

$$X_i(t) = S_i(t) + N_i(t)$$

where $S_i(t)$ is the signal (perfectly correlated across channels) and $N_i(t)$ is the (uncorrelated) noise. Then the signals are proportional to one another:

$$S_i(t) = a_i\, f(t).$$

We accept the following as the best estimate for the relative amplitude:

$$\langle a_{i,est}\rangle = h_i \sum_{j\neq i} C_{ij},$$

where the normalizing factor is $h_i = (\sum_{j\neq i} a_j)^{-1}$. Then it is straightforward to show that the estimate is correct:

$$\langle a_{i,est}\rangle = a_i.$$

The variance in the estimate is:

$$VAR(a_{i,est}) = \langle a^2_{i,est}\rangle - \langle a_{i,est}\rangle^2 = \varepsilon^2 [\ 1 + 1/(N-1) + T\varepsilon^2/(N-1)a^2\ ]$$

where a is a typical value for the a_i, and ε^2 is the variance of the coefficients describing an expansion of the noise in terms of orthonormal functions. Consequently, for small noise levels ε and for a large number N of samples in the recording, $VAR(a_{i,est}) = \varepsilon^2$.

To understand the implication of this, consider a 1-sec epoch in which sampling was obtained at 128 Hz (i.e., N = 128). The covariance procedure produces a single number from the 128 samples that are used to compute each of the C_{ij}, and in this way reduces the noise level. More precisely, it reduces the noise power by 128, or the rms noise by a factor of 11. Additional noise is introduced by the absence of a perfect reference signal, but this is a minor addition to the noise. Herein lies the advantage of applying covariance techniques to improve the estimates for the relative signal strengths across a multi-sensor system. This approach greatly enhances the precision with which an equivalent dipole source can be located.

2.10. Probe Position Indicator

We have purchased and received a Probe Position Indicator (PPI) system, which was described in our original proposal. An identical system is installed at NYU's Bellevue facility, and members of our group have already conducted a full evaluation of it. Therefore, we know that this system will give us head position relative to neuromagnetometer sesnors with high accuracy. Although the PPI was delivered, it is not yet fully installed. A report will be submitted after installation and testing.

3. Color Vision Studies

Many people have speculated that the mechanisms underlying the perception of chromatic variations in light, as opposed to luminance variations, are located at different places in the cortex. To test this idea, Krauskopf, Kaufman and Williamson are preparing to record magnetic fields generated when an observer views a field modulated only in color or only in luminance. For this purpose we have purchased an Adage 3000 frame buffer and a suitable TV monitor. To date we have written some programs to calibrate the system and generate the sorts of stimuli needed for the experiment.

We anticipate starting this experimental work in late November, shortly after David Travis, a new post-doctoral fellow arrives.

4. Memory Studies

As indicated in our original proposal, we planned to conduct a series of studies combining the methods of neuromagnetism with psychological methods for the study of memory and related cognitive processes. Toward this end we developed computer programs that would allow us to present verbal stimuli on the video screen of an Amiga computer. The words were selected to be either highly imageable or highly abstract, since memory for imageable (concrete) words is better than it is for abstract words. In addition we recently acquired a video projector that will repeat the image on the Amiga screen and allow us to project it into the magnetically shielded room without contaminating the magnetic environment with unwanted noise. We have also fully worked out "hanshaking" software so that the main HP 900-550 computer will "know" which word is presented, and sort responses into vsarious categories for subsequent analysis.

While all of this work was going on, we held a seminar in which we reviewed relevant papers on memory, psychophysiology and mental imagery as well as on effects of brain damage on these processes. Drs. Glanzer, Kaufman, Schwartz participated in this seminar, along with one graduate student who was assigned to the project. Unfortunately, this student had to withdraw from the program because of personal problems. We will replace her with another student, and have already enlisted the help of Dr. Eliot Hershman, a post-doctoral fellow in the department of Psychology.

While all of this backgrou nd work was going on, Drs. Glanzer and Adams (another post-doctoral fellow) were able to set up a simple task for the reliable demonstration of the effect of the difference between imageable words and abstract words in memory. In a second experiment using the same materials they showed that the concreteness effect can be increased by requiring that the subjects try to evaluate the concreteness of each word. (Evaluation of imageability and concreteness give highly correlated judgements). The effect of this procedure is an increase in memorability of the items. One hypothesis that we will test in the neuromagnetic studies is that the act of evaluating imageability enhances activity of the areas of the cortex involved in vision. We shall test this hypothesis using visual presented words, and also spoken words.

The work alluded to above is now in preparation. As soon as completed manuscripts are available, they will be transmitted to AFOSR.

It should also be noted that we upgraded the memory research facilities available to Dr. Glanzer. This was intended to replace his Apple computer based laboratory with an IBM AT compatible facility. Also, a duplicate facility was provided for Dr. Kaufman in the interest of sharing capabilities for joint work.

5. The Dimensional Structure of Knowledge Spaces

An important area of investigation of Falmagne's research group is the development of computerized procedures to assess, in an economical way, the state of knowledge of individuals in some - specified, but arbitrary - domain of information. This research is based on the principle that a student's knowledge state can be formalized as the subset of all the questions (problems) that this individual is capable of solving. Not all subsets of questions are possible knowledge states. For instance, in the field of thus, a subset of problems containing one involving long division, but not containing one involving subtraction, is not a possible knowledge state. For each domain of knowledge, represented here as a (large) collection of questions or problems, there is a particular family of subsets of questions, called the *knowledge structure*, that contains all possible states of knowledge in that domain.

The part of this broad area of research that is relevant here is the analysis of the dimensional aspects of a knowledge structure. Even though this may not be obvious at first, the concept of knowledge structure is consistent with a multidimensional representation. This is of interest, since such a representation may be suggestive of underlying abilities explaining the subjects performance. A comparison with more traditional psychometric methods (*tailored testing*) becomes then possible. However, there are several ways to impose a multidimensional representation on a knowledge structure (see Doignon and Falmagne, 1985). In our view, the potentially most fruitful one is based on the notion of *biorder dimension*, as defined by Doignon, Ducamp and Falmagne (1984) in their mathematical analysis of multidimensional Guttman scales. In that context, a relation %R%, with %sRq% representing a correct response by subject %s% to question %q%, is analysed as the intersection of a (minimal) number of special relations:

$$R = \mathop{IN}_{i=1}^{n} B_i ,$$

where each %B sub i% is a *biorder* relation, symbolizing the concept of a (unidimensional) Guttman scale. Concretely, this means that each %B sub i% satisfies

$$s\, B_i\, q,\ not\ s'B_i\, q,\ s'B_i\, q' \implies s\, B_i\, q'.$$

The biorder dimension of %R% is the minimal number %n% for which such a representation exists. In our context of knowledge structures, a subject gives a correct response to a question if and only if the question is contained in the subject's knowledge state. This means that we can apply the analysis of multidimensional Guttman scales to the membership relation between questions and knowledge states. In doing so, we define the dimension of a knowledge structure as the biorder dimension of its associated membership relation.

Although we are dealing here with problems that are NP-complete (cf. Doignon, Ducamp and Falmagne, 1984), procedures for actually solving this biorder representation problem -- that is, for computing the bidimension and for constructing one or more suitable representations -- have been developed (Chubb, 1986; Koppen, 1987). In view of these results, it seems reasonable to conjecture that no unsurmountable difficulties will arise in practical applications. It must be noted, however, that these solutions of our representation problem all pertain to the deterministic case, where we want the representation of the relation as the intersection of a number of biorders to be perfect. In practice, errors of various sorts will be present in the typical data. Extending this approach to probabilistic situations is the topic of the current research. The research team consists in Jean-Claude Falmagne (PI; one summer month per year) and Mathieu Koppen (2/3 of time), a research scientist who joined the program in March 1987. Since, to the best of our knowledge, the problem of defining probabilistic models for the multidimensional Guttman scale is new, and in this form barely touched upon in the literature, we devoted the first 6 months of our research to a systematic exploration of different directions that seem promising in this respect.

We started by developing the following non parametric model for %P(s,q)%, the *probability* that subject %s% answers correctly to question %q%:

$$P(s,q) = \begin{cases} aa_s & \text{if } s\,A\,q, \\ bb_q & \text{otherwise,} \end{cases}$$

where %A% is an approximate, low dimensional biorder representation of the observed relation. That is, %A% is the intersection of a small (%n"="2% or %n"="3%, say) number of biorders that only approximately reproduce the observed relation. The %aa sub s% and %bb sub q% are parameters that depend on the subject and the question, respectively, and have values in the real interval %[0,1]%. We have for this model the following interpretation. The low dimensional approximation %A% represents the "real", latent state of affairs and the probability of *observing* a correct response by subject %s% to question %q% depends on his latent state, and on some chance factors. In particular, we suppose that if a subject %s% has mastered question %q% (%sAq%) there is a probability %aa sub s% of observing a correct response; in the other case, if a subject %s% has not mastered some question %q%, he still might give a correct reponse with a probability %bb sub q%. This means that %1- aa sub s% can be interpreted as a "careless error" parameter associated with subject %s%, and %bb sub q% as a "lucky guess" parameter associated with question %q%. Clearly the deterministic case (no error) corresponds to the situation where %A% represents the observed relation perfectly, %aa sub s^=^1% for all %s% and %bb sub q^=^0% for all %q%. We derived the likelihood equations for this model in terms of the parameters %A%, %aa sub s% and %bb sub q%. It appeared that these equations could be readily solved for the %aa% and %bb% parameters in terms of %A%. So we are left with the problem of solving the likelihood equations as a function of the %A% parameter only. Here %A% ranges over the collection of all approximate biorder representations of the observed relation having some fixed (low) dimension. There appear to be no numerical problems involved in finding the maximum likelihood estimate, but the problems are rather combinatorial, due to the multitude of possible approximations and the discrete nature of this parameter. This precludes, for the moment, the use of standard maximum likelihood techniques. The problem can be seen as that of defining a suitable topology on the collection of approximations under consideration.

Another approach that we are investigating is based on embedding our problem in the wide class of probabilistic models known as *random utility* models (see, e.g., Block and Marschak, 1960). Traditionally, these models are applied in econometrics and psychology in the context of choice theory, but their use is certainly not restricted to preference data. In the case of binary choice data these models attack the problem of representing a collection of observed choice probabilities %P sub ab% of choosing %a% over %b%, where %a% and %b% run through some fixed set %X% of alternatives, by a collection %"{"^bold U sub a":"a mo X^"}"% of jointly distributed random variables, such that

$$P_{ab} = Pr(\mathbf{U}_a \geq \mathbf{U}_b).$$

In our case the relation under consideration is not a preference relation, but it still is a dominance relation, which allows us to use the same approach. That is, we can formulate the problem of finding collections %"{"^bold U sub s~:~s mo S^"}"% and %"{"^bold V sub q~:~q mo Q^"}"% of random variables, such that

$$P(s,q) = Pr(\mathbf{U}_s \geq \mathbf{V}_q).$$

In this way the unidimensional Guttman scale problem is defined as an instance of the random utility model, apart from the fact that we are dealing with two distinct sets (subjects and questions), rather than one set of alternatives and thus define two collections of random variables instead of one. The random variable %bold U sub s% can readily be interpreted as representing the ability of subject %s%; likewise, the random variable %bold V sub q% represents the difficulty of question %q%. So the model simply says that the probability that subject %s% solves question %q% is equal to the probability that the ability of subject %s% exceeds the difficulty of question %q%. This corresponds exactly to the classic interpretation of the Guttman scale, with the only distinction that here ability and difficulty are seen as random variables, not as fixed values. Note that this model generalizes the standard unidimensional psychometric models. The extension to the multidimensional case is now straightforward. For a representation in %n% dimensions we search for collections %"{"^bold U sub {s,i}~:~s mo S,~i=1...n^"}"% and %"{"^bold V sub {q,i}~:~q mo Q,~i=1...n^"}"% of random variables, such that

$$P(s,q) = Pr(\mathbf{U}_{s,i} \geq \mathbf{V}_{q,i} \text{ for } i=1 \cdots n).$$

We are investigating to which extent this approach allows us to use results obtained in the area of random utility models for the representation of probabilities that derive from a multidimensional Guttman scale.

6. Activities in the Visual Neuroscience Laboratory

The contribution of the URI to Movshon's laboratory in the present budget year was to assist in the acquisition of two processor upgrades for the PDP11/34 computer systems controlling the neurophysiology and human psychophysics laboratories. The new processors (Nissho model 1100 and 1100+, PDP 11/84 equivalent) have more than doubled the performance of these systems. The unit in the neurophysiology laboratory, a model 1100+, is equipped with 4 Mbytes of main memory and 16 kbytes of cache memory; in the human psychophysics laboratory, a model 1100 with 2 Mbytes of main memory is in use. The rationale behind the purchase of these units was to improve the capacity of the laboratory computer systems, especially for the generation of dynamic stochastic random motion displays of the kind used in studies of Schwartz, Kaufman and Movshon on the perception of changing direction of motion. This work has been submitted as a poster to the forthcoming Neuroscience meeting, but owing to existing vagaries of funding, we may not be able to be reepresented at the conference in November. However, as soon as a manuscript is available, it will be transmitted to AFOSR.

These displays consist of a stream of randomly positioned dots, within which a small proportion of nonrandom dots carry an unambiguous motion signal. They have the unique virtue of testing sensitivity to *motion* while leaving factors such as spatio-temporal contrast energy constant, and thus can serve as specific probes for the function of the visual motion system. The machines purchased with URI funds are now in use in the psychophysical and electrophysiological testing of motion sensitivity using these displays.

The *physiological* testing (in collaboration with Drs. W. T. Newsome and K. Britten at the State University of New York at Stony Brook) has concentrated on the properties of neurons in the extrastriate visual area MT, which is known to have important functions in motion analysis. This work has revealed that the sensitivity of MT units to the motion signal in the stochastic displays is often as good as or better than the sensitivity of trained primate observers performing a behavioral motion discrimination. We expect to find that the sensitivity of units earlier in the motion pathway is less good, largely because these units cannot integrate signals over space as effectively as MT units.

The *psychophysical* testing here at NYU has concentrated on measuring the spatial and temporal determinants of sensitivity to motion in these displays. Although the results are incomplete at this writing, the data clearly reveal that the system integrates over considerably larger regions of both space and time in analyzing these patterns than it does in analyzing simple contrast patterns. We are in the process of developing an ideal-discriminator model in order to see how the performance of observers compares with the performance of devices that make maximal use of the available information in the display.

As soon as the new video projector system is installed, we also plan human studies using the neuromagnetometer to identify areas of human cortex that are analogous to MT in rhesus macaque.

7. Visual Learning and Measurement Research

During the first year of the URI program at NYU, Dr. Snodgrass focussed on two areas -- new and continuing research on forms of visual learning, as studied with the picture fragment completion task, and completion of some research on measurement of discrimination and response bias in recognition memory of normal and abnormal populations. The visual learning research is described in two published or in-press papers and one submitted paper (see publication list), and the measurement paper is in press in the Journal of Experimental Psychology: General. The visual learning research was or is scheduled to be presented at three conferences, and the measurement research was presented at the international meeting of the Neuropsychology Society.

7.1. Visual Learning

The general paradigm used in this research consists of two or three phases. In Phase 1, subects are presented with a series of fragmented pictures on the display of an Apple Macintosh computer and asked to identify them. In the full series presentation (the "Basic Task"), pictures are shown in an ascending order of fragmentationfrom most fragmented (Level 1) to least fragmented or complete (Level 8). At each level they attempt to identify the picture by typing its name. If the typed name matches one of the correct names stored for each picture, the series is terminated and the threshold for that item is recorded as the level of fragmentation at which it was correctly identified. In a single level presentation (also called the Priming Task), subjects are shown pictures at a single level of fragmentation (usually one of the levels 3, 5, or 8), and asked to identify it. They are informed whether their response is correct or incorrect, and also given the correct name. The present research, unlike prior work where we used 18 pictures during training, we use 33 pictures (3 for practice and 30 for experimental). A 10-minute interval filled with a distractor task intervenes between Phase 1 and Phase 2 to eliminate recency effects in memory for training figures. In an ongoing experiment we are manipulating this inter-Phase interval over the range of 10 minutes to two weeks.

In Phase 2 we test for the subject's perceptual learning of the items presented during Phase 1. This is defined as the degree of savings achieved on items repeated from Phase 1 (old items) compared to baseline performance achieved on a comparable group of ne or never-before-seen items. The full series of pictures, using the ascending method of limits of Phase 1 of the Basic Task, is presented in Phase 2. In ongoing research we are using 60 pictures rather than the 30 pictures of Phase 2. These are divided into 30 old and 30 new.

In Phase 3 we test subjects for episodic learning of items presented during Phases 1 and 2. This usually takes the form of a computerized recall test, in which subjects are asked to type the names of the pictures from either phase of the experiment. Responses are scored for correctness using the same files of correct names used to score identifications in Phases 1 and 2. We have also implemented a *yes/no recognition test for this phase.*

The basic question being asked with this paradigm is: What kinds of training maximize perceptual learning, what kinds of training maximize episodic learning, and are these the same variables? Since perceptual learning can be exhibited in the absence of explicit memory for the prior occurrence of an item, are the two kinds of memory different? We must now consider some methodological issues.

7.1.1. Item Effects

A striking characteristic of picture fragmentation completion performance is the variability in difficulty from item to item. This variability is greater than subject-to-subject variability in the case of populations of young subjects. Because perceptual learning is measured by comparing performance on the old item set with that on the new item set, item variability is a potential problem. While standard counterbalancing schemes tend to average out effects of this variability, a more powerful method is to equate old-new sets on the basis of perceptual identification performance. The results of the normative experiment described in Snodgrass, Smith, Feenan, and Corwin (1987) were used to construct new sets equated on the basis of performance, and these are being used in our ongoing research.

7.1.2. Measurement of Savings

In our earlier work we measured savings in perceptual learning by subtracting old from new thresholds, in accordance with common practice.

However, this assumes a linear learning function which, in the limit, is wrong. Accordingly, we are now testing various learning functions against our normative data base. The results of this work will inform our future procedures.

7.2. Results to Date

To date, we have carried out four experiments in which training was manipulated and the effects on perceptual learning and, in two experiments, episodic learning was measured. Using the basic task, subjects were presented with one of four conditions during Phase 1: (1) the full series without presentation of the full picture after identification; (2) the full series with the full picture; (3) only the full picture; and (4) the full series with different fragments from Phase 2. Perceptual learning was large (about 2 levels) and equivalent for conditions (1) and (2); signficantly less (about 1 1/2 levels) for condition (4), and least of all (less than 1 level) for condition (3).

The lack of a difference for conditions (1) and (2) was surprising, as we thought that to the extent that explicit memory for the picture helps performance, subjects would do better when shown the full picture. Although showing only the full picture in condition (3) produced significant perceptual learning (this is the standard repetition priming paradigm), it produced the least perceptual learning of all conditions, and suggest that subjects benefit most from seeing some fragments, even if they are not the same fragments as shown at test, as in condition (4).

These results do not, however, tell us to what extent subjects are doing the task either by explicit recognition of a repeated fragment or by lucky guesses from a pool of training pictures. Future work will permit us to investigate such questions.

The relationship between recall and perceptual learning was explored in the different fragments experiment and in a fourth experiment which presented only one of three levels of fragmentation during training (using the priming paradigm for Phase 1). The patterns of relationship are complex and difficult to summarize in the available space. Suffice it to say that somnetimes perceptual and episodic memory are associated, and sometimes they are dissasociated by being either unrelated or negatively related.

7.2.1. Measurement

The work in measurement concerns the question of how to measure discrimination and bias in recognition memory so that the two measures are theoretically independent. In the paper by Snodgrass and Corwin (in press), we analyzed seven models of recognition memory data in which both discrimination and bias were manipulated. Of the seven models, we found that only a signal detection theory model with the response bias measure C (distance from the intersection of the old and new distributions) and a two-high-threshold model satisfy theoretical and pragmatic independence. Discrimination and bias measures from these two models comparing recognition memory of patients with dementing diseases, amnesics, and normal controls. We found the two-high-threshold model to be more sensiitive to group differences, but the SDT model was more plausible as a theoretical explanation for the results.

8. Support of Facilities

All of the collaborators on this project communicate with a department owned VAX computer known colloquially as 'xp'. This report is stored in xp and is printed out on its Apple Laser Writer. Number crunching, report writing, and visual displays for some classes of perception research are done through xp. Therefore, the modest sum of $1000 was expended as a contribution to the maintenance contract on xp.

Similarly, Dr. Falmagne is conducting his research into the structure of knowledge using another VAX computer known as 'Aris'. This work complements other work he and his colleagues are conducting in the field of automated instruction, which in itself depends upon a theory of how to test knowledge structures. Hence, another $1000 was expended to support service for his VAX computer.

About $4000 was expended to service the HP 9000 550 computer which is at the heart of the upgraded neuromagnetism facility. This is a major facility improvement which was paid for out of several different grants. However, in view of the central importance of the new Neuromagnetism Laboratory (NML) to the URI program, it is important that it be fully supported.

9. Training

As indicated in the proposal, this program has a strong and important training component. Owing to cost limitations, we were unable during the first year to bring in post-doctoral trainees in many areas. However, we did provide partial support for Dr. Risto Ilmoniemi who has already developed an entirely new way in which to analyze and interpret neuromagnetic fields. This work is described above.

Two graduate students in psychology and two in physics have received either full or partial support under the aegis of the URI. During the next year we anticipate being able to support more students and post-doctoral scientists as the need for major capital improvements will not be nearly so demanding.

Several ongoing seminars were conducted. These ranged from cognition to brain and behavior, and from sensory and perceptual processes to computer vision and mathematics. The anoucments of these seminars were widely dispersed and they were well-attended. A listing will be provided on request.

10. Interaction with DoD Laboratories

Most of our interactions have been with the Aerospace Medical Research Laboratory at WPAFB. Kaufman made two visits and Dr. Glenn Wilson and Captain Paul Drago made visits to NYU. We are currently collaborating on an experiment to determine if a well-known effect of ISI on auditory evoked fields also occurs in the case of visual evoked fields. In addition, Kaufman has assisted Dr. K. Boff by reviewing the glossary for his forthcoming Compendium. This was done without compensation. Also, Williamson and Kaufman have reviewed proposals for AFOSR. It is possible that Sperling also reviewed such proposals, but the author of the present document has no direct knowledge of this. Also, Kaufman delivered the keynote address at the forthcoming NATO AGARD meeting in Trondheim, Norway. This address describes many of the activities at NYU as well as the facilities made possible by the award of the URI.

11. **Travel** Two of us attended a users conference at BTi in San Diego in January of this year. This was at no cost to the Air Force, although our project benefited from attendance at the conference. Kaufman attended the Neuroscience conference in Budapest and a sattelite conference in Graz, Austria on Magnetoencephalography, where we he presented a paper on work at the NYU Neuromagnetism Laboratory. Finally, two of us attended the World Biomagnetism Conference in Japan in September 1987, and a paper was presented there too.

12. References

Block, H.D. & Marschak, J. (1960). Random orderings and stochastic theories of responses. In Olkin, I., Ghurye, S., Hoeffding, W., Madow, W. & Mann, H. (Eds.), *Contributions to probability and statistics* (pp. 97-132). Stanford, CA: Stanford University Press.

D.S. Buchanan, D. Paulson, and S.J. Williamson (1987). Instrumentation for Clinical Applications of Neuromagnetism, In: *Advances in Cryogenic Engineering*, R.W. Fast, Ed. (Plenum Press, New York), in press.

J. Calvet, M.C. Calvet, and J. Scherrer (1964). Etude stratigraphique de l'activité EEG spontanée. *Electroenceph. clin. Neurophysiol.* 17, 109-125.

P. Carelli, V. Foglietti, I. Modena, and G.L. Romani (1983). Magnetic study of the spontaneous brain activity of normal subjects. *Il Nuovo Cimento* 2D, 538-546.

G.E. Chatrian, L. Bergamini, M. Dondey, D.W. Klass, M. Lennox-Buchthal, and I. Petersén (1974). A glossary of terms commonly used by clinical electroencephalographers. *Electroenceph. clin. Neurophysiol.* 37, 538-548.

Chubb, C. (1986). Collapsing binary data for algebraic multidimensional representation. *Journal of Mathematical Psychology*, **30**, 161-187.

D. Cohen, E. Edelsack, and J.E. Zimmerman (1970). Magnetocardiograms taken inside a shielded room with a superconducting point-contact magnetometer, *Appl. Phys. Lett.* 16: 278-280.

Creutzfeldt, O., and Houchin, J. (1974). Neuronal basis of EEG-waves. In: Rémond, A., *Handbook of Electroencephalography and Clinical Neurophysiology, Vol. 2, Part C, The Electrical Activity from the Neuron to the EEG and EMG, The Neuronal Generation of the EEG,* Elsevier, Amsterdam, pp. 5 - 55.

Doignon, J.-P., Ducamp, A. & Falmagne, J.-C. (1984). On realizable biorders and the biorder dimension of a relation. *Journal of Mathematical Psychology*, **28**, 73-109.

Doignon, J.-P. & Falmagne, J.-Cl. (1985). Parametrization of knowledge structures. Mathematical Studies in Perception and Cognition, 85-1. Department of Psychology, New York University. Submitted.

I.D. Frost (1968). EEG-intracellular potential relationships in isolated cerebral cortex. *Electroenceph. clin. Neurophysiol.* 24, 434-443.

M.S. Hämäläinen, and R.J. Ilmoniemi (1984). Interpreting measured magnetic fields of the brain: estimates of current distributions. Report TKK-F-A559, Helsinki University of Technology.

R.J. Ilmoniemi, S.J. Williamson, and W.E. Hostetler (1988). New Method for the Study of Spontaneous Activity, to appear in: *Biomagnetism '87: Proceedings of the 6th International Conference on Biomagnetism*, Tokyo Denki University Press, Tokyo), in press.

V.O. Kelhä, J. M. Pukki, R. S. Peltonen, A. J. Penttinen, R. J. Ilmoniemi, and J. J. Heino (1982). Design, construction, and performance of a large-volume magnetic shield, *IEEE Trans. Magn. MAG* 18:260.

Koppen, M.G.M. (1987). On finding the bidimension of a relation. *Journal of Mathematical Psychology*, **31**, 155-178.

Lopes da Silva, F.H., and Storm van Leeuwen, W. (1978). The cortical alpha rhythm in dog: the depth and surface profile and phase. In: Brazier, M.A.B., and Petsche, H., *Architecture of Cerebral Cortex*, Raven Press, New York, pp. 319 - 333.

G. L. Romani and L. Narici (1986). Principles and clinical validity of the biomagnetic method, Med. Prog. thru. Tech. 11:123.

Vvedensky, V.L., Ilmoniemi, R.J., and Kajola, M.S. (1986). Study of the alpha rhythm with a 4-channel SQUID magnetometer. *Med. & Biol. Eng. & Computing* 23, Suppl. Part 2, 11-12.

A. Mager (1981). The Berlin magnetically shielded room (BMSR), Section A: Design and construction, In: *Biomagnetism*, S.N. Erné, H. D. Hahlbohm, and H. Lübbig, Eds. (Walter de Gruyter, Berlin), p. 51.

Y. Okada, L. Kaufman, and S. J. Williamson (1982). Hippocampal formation as a source of endogeneous slow potentials, *Electroenceph. clin. Neurophysiol.* 55:417.

S. J. Williamson, M. Pelizzone, Y. Okada, L. Kaufman, D. B. Crum, and J. R. Marsden (1984). Magnetoencephalography with an array of SQUID sensors, In: *Proceedings of the 10th International Cryogenic Conference - ICEC10* (Butterworth, London), p. 339.

J. E. Zimmerman (1977). SQUID instruments and shielding for low-level magnetic measurements, *J. Appl. Phys.* 48:702.

NEW METHOD FOR THE STUDY OF SPONTANEOUS BRAIN ACTIVITY

AFOSR-TR- 87-1710

R.J. Ilmoniemi, S.J. Williamson, and W.E. Hostetler

Neuromagnetism Laboratory, Departments of Physics and Psychology
New York University, 4 Washington Place, New York, NY 10003, U.S.A.
and
Center for Neuromagnetism, Department of Physiology and Biophysics
New York University Medical Center, 550 First Avenue, New York, NY 10016, U.S.A.

INTRODUCTION

We have developed new methods to analyze multichannel neuromagnetic recordings of spontaneous brain activity that avoid specific assumptions concerning the nature of the sources. This approach can be applied to studies of a variety of brain signals, such as alpha rhythm and interictal epileptic activity, for the purposes of classification and analysis, even if the field pattern over the scalp may not be characterized as that of a single current dipole. We illustrate this approach by applying these methods to the study of the alpha rhythm in human subjects.

By definition (Chatrian et al., 1974), the alpha rhythm is brain activity that gives rise to electrical oscillations between 8 and 13 Hz on the occipital scalp and is attenuated by visual stimuli. The cortical origin of alpha activity has been evidenced by studies of potentials at various depths within the cortex of animals (Calvet et al. 1964; Creutzfeldt and Houchin, 1974; Frost, 1968) and especially by a clear polarity reversal between superficial and depth electrodes in dog (Lopes da Silva and van Leeuwen, 1978). These last studies also provided evidence that alpha activity originates in different epicenters from which activity spreads in several directions, rather than originating in a single source and sweeping over a large area of cortex.

Previous neuromagnetic studies (Carelli et al., 1983; Vvedensky et al., 1986) suggest that many sources responsible for magnetically monitored alpha activity are located near or in the visual cortex and there are time series of the rhythm during periods of strong activity in which the oscillation period is constant. These time series are called *spindles*. The magnetic field pattern during a spindle appears to remain relatively stable. Based on these findings, we adopt the working hypothesis that there are *specific configurations of neurons* that exhibit such oscillatory excitations. We call such oscillations *alphons*. These hypothetical oscillations would be basic units of activity, whose signals add up to form the observed alpha rhythm. Our analysis was developed to test this alphon hypothesis by determining whether the underlying sources of spindles can be said to differ significantly. In particular it would be interesting to determine whether an alphon can be associated with many spindles, indicating that the alphon exhibited repeated oscillations.

It is possible to characterize the source configuration of an alphon without the use of any model, such as a current dipole model. We study the kinematics of the alpha rhythm in an n-dimensional "signal" space, as defined by the output amplitudes of the n-sensor system. The lead fields of n detection coils are not generally orthogonal in current space (Hämäläinen and Ilmoniemi, 1984) nor do they form a complete set of basis vectors (they span a subspace of the current space). Nevertheless, even without orthogonalizing the set of n basis vectors we may investigate the distribution of activity within this space. A given source configuration corresponds to a specific direction for its *spindle vector* in signal space, as defined by the 14 sensor outputs, and a greater distance from the origin in a fixed direction represents greater source strength. Thus we may ignore source strength *per se* in our discussion if we confine consideration to direction alone. For instance, we can determine whether more than one spindle can be attributed to the same alphon by establishing whether the spindle vector in this space points in the same direction as the spindle from that alphon (the *alphon vector*). Furthermore, it would be possible to determine whether different directions of signal space have characteristic features, such as different alpha frequencies.

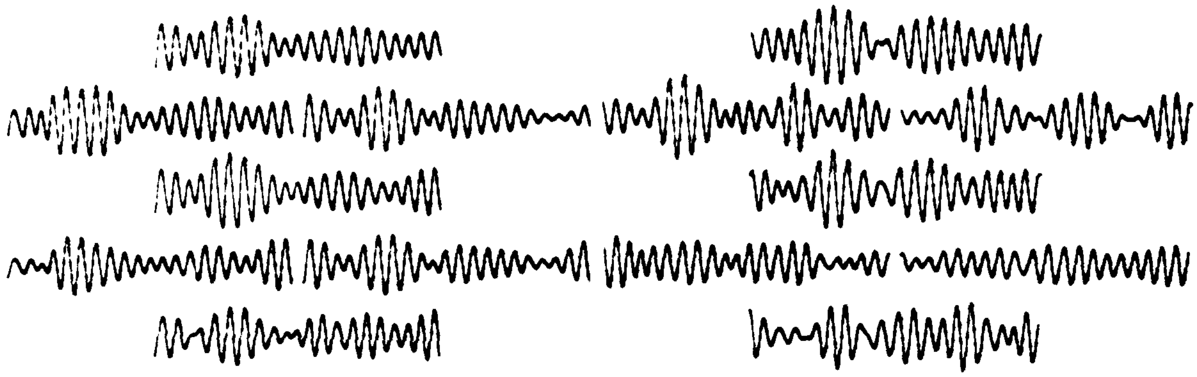


Fig. 1. Observed 2-second time series from subject RJI within the bandwidth from 8 to 13 Hz showing alpha activity. The array of recordings indicates the relationship of the 7 detection coils within an individual probe, but the relative orientation of one probe to another is not correctly indicated.

METHODS

Alpha activity for two subjects was recorded with the GEMINI system at the Center for Neuromagnetism of the New York University Medical Center. This system has two probes in separate dewars, each with seven second-order SQUID gradiometers. When maintained in fixed positions the system defines a 14-dimensional signal space. The geometry of the detection coils within each probe is identical to that of a 5-sensor probe described by Williamson et al. (1984) except that each central detection coil is surrounded by 6 rather than 4 outrider coils, which lie on a circle 4 cm in diameter. The subject was prone on a bed in a room with subdued lighting, with an unobstructed downward view onto a patterned surface. The probes were positioned over the occipital area of the scalp about 5 cm above the inion and 6 cm to either side of the midline, where the alpha signals were found to be strongest (Fig 1a). Data were recorded continuously for 12-second epochs, with the subject's eyes closed; and at fixed intervals recordings were done with eyes open to verify the existence of alpha blocking in this condition and to obtain an estimate of the background noise. The recording bandwidth was 1-50 Hz; for analysis, the passband was narrowed to the alpha band, i.e., 8-13 Hz.

The initial analysis of the data consists of using a computer to automatically detect intervals of strong spindles and picking those that have a stable period and no phase shifts across sensors. Typically, 1 to 5 such spindles were detected during a 12-second epoch. If the alphon hypothesis were valid, these spindles could be due to individual alphons, and the field pattern should remain stable in signal space during each spindle's lifetime. Figure 1 shows representative data for the 14 sensors. To improve the signal-to-noise ratio in determining spindle amplitudes, the covariance between the signal of a given sensor and each of the other signals was computed. These covariances were added together and divided by the sum of the other signal amplitudes to obtain the portion of the given sensor's signal that is coherent across sensors. The 14 amplitudes thus obtained define the components of the spindle vector in signal space.

To determine whether the source configuration of two spindles differ significantly, we need only determine whether the angle between their spindle vectors differs significantly from the noise. Conceptually, it may be useful to consider noise for each measurement as represented by an ellipsoid in signal space (Fig. 2a). It is an ellipsoid rather than a sphere, since generally the instrumental, environmental, and subject noise differ across sensors, and the latter two may also be correlated across sensors. If we imagine such a noise ellipsoid centered on the head of each spindle vector, then when the ellipsoids of two spindle vectors do not have angular overlap, the measurement has indicated the existence of separate source configurations.

All pairs (i,j) of spindle vectors were analyzed to determine whether their angle of separation α_{ij} significantly exceeds the noise. We determined this by projecting in turn the background noise for each of the 14 sensors onto the plane defined by the pair of spindle vectors, and adding these noise vectors to the ith spindle vector with signs for their components that most quickly bring the resultant vector toward the jth spindle vector (Fig. 2b). The corresponding angle between the ith spindle vector and the resultant when noise is added is called the *noise angle* η_{ij}. In a similar fashion the noise angle η_{ji} for the second spindle vector is

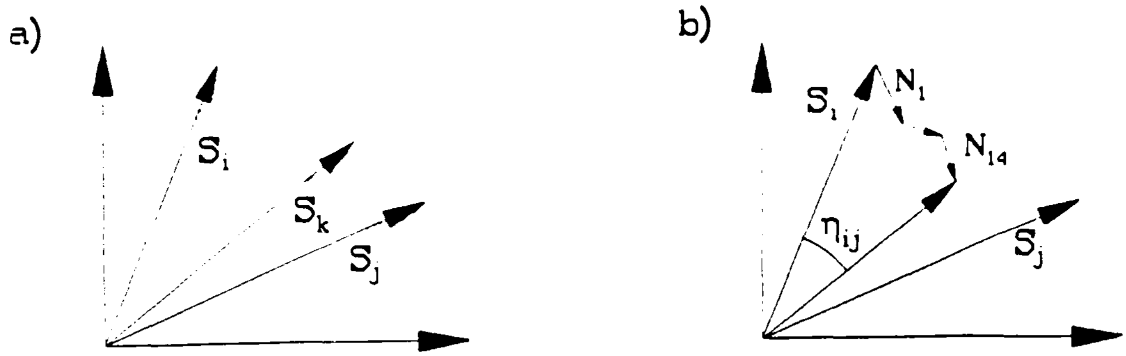


Fig. 2. (a) A representation of spindle vectors S_i, S_j, S_k and the noise ellipsoids in 14-dimensional signal space. The angle between S_i and S_j is denoted by α_{ij}. (b) Definition of the noise angle η_{ij} from S_i toward S_j, obtained by adding projected noise vectors $N_1, N_2, \ldots, N_{14}$ for each of the sensors.

computed. The total noise angle is then taken as $\varepsilon_{ij} = [\eta_{ij}^2 + \eta_{ji}^2]^{1/2}$. We introduce the *discrimination ratio* $D_{ij} = \alpha_{ij}/\varepsilon_{ij}$ to represent the angular separation of two spindle vectors compared to the noise. D_{ij} is a measure of the significance of the difference between a pair of spindle sources. All spindles that cannot be distinguished from a given one are called its *neighbors*.

The data for 4 trials with subject RJI and 3 trials for subject FAL were analyzed as described above. Each trial contained 18 twelve-second epochs, the epochs being separated by 8 seconds. In each trial, 26 to 35 spindles were found, based on the criteria that a spindle has to exceed a minimum amplitude of 0.5 pT and a minimum duration of 500 ms, and that the oscillations in different sensors must be coherent. In addition, phase stability within each channel during a spindle was required.

RESULTS

Typical spindles had durations under 2 sec and peak-to-peak amplitudes of about 1-2 pT. The range of observed frequencies for RJI was from 10.0 to 11.2 Hz and for FAL from 9.6 to 10.8 Hz. Typical angles between spindle vectors in signal space were 10-30 deg. With our signal-to-noise ratio of about 8, we found for both subjects that almost every spindle could be distinguished from each of the others in a trial with a discrimination ratio $D_{ij} = 2.0$. Therefore, each spindle characterizes a different alphon at this discrimination level. The same was true with $D_{ij} = 4.0$ for FAL, but at this level for RJI only 9-12 alphons were needed to explain each set of 27 to 35 spindles in a trial. Using the terminology defined in Fig. 3, a typical alphon could account for 3 to 10 spindles (the *cohorts* of the alphon), and each spindle could be explained by any of 1 to 6 alphons (*candidates* for the spindle). Successive spindles were often less distinguishable from each other than spindles separated by longer times. Generally, as the signal-to-noise ratio increases so that spindles may be

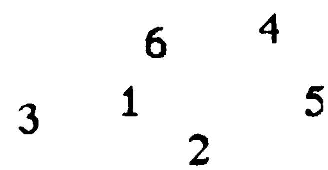


Fig. 3. Illustration of spindle relationships in signal space. The neighborhood of spindle vector 1 (stipled) includes spindles 2, 3 and 6, so these three spindles are its neighbors. If spindle 1 defines an alphon, spindles 2, 3 and 6 are its cohorts. The neighborhood of spindle 4 includes spindles 5 and 6, so if spindle 4 defines an alphon these two are its cohorts. From the point of view of spindles that do not define alphons, spindles 2, 3 and 6 have alphon 1 as their candidate, and spindles 5 and 6 have alphon 4 as their candidate. Consequently, spindle 6 has two alphons (1 and 4) as candidates for its source.

better discriminated from each other, we expect to have fewer cohorts for each alphon and fewer candidates for each spindle.

Spindles that are cohorts of a given alphon were often found to have differing frequencies. Therefore, either an alphon is not limited to oscillating at a given frequency each time it is excited, or with improved signal-to-noise ratio we could be able to distinguish between these spindles of differing frequency.

CONCLUSIONS

This study shows it is possible to distinguish between most of the sources of observed spindles using the spindle vector representation. For the present subjects, some spindles that are cohorts of an alphon are found to have differing physical properties such as frequency. This implies that to within the resolution of our measurements the underlying neural excitation with fixed geometry can be modulated. Our results indicate that the alpha rhythm is generated by a large number, or possibly a continuum, of different source configurations. It remains to be seen whether most or even any of these are locally oscillating portions of cortex so that they might be modeled by current dipoles.

It has not escaped our attention that analyses of this type can be applied to classify the sources of signals from other types of neural excitations, such as interictal epileptic activity. One advantage is that this method is not computationally demanding, and model-specific analyses need be performed only once for each class, such as for an alphon representing all of its cohorts.

ACKNOWLEDGEMENTS

We thank L. Kaufman, L. Cornacchia, F. Lado, and Jia-Zhu Wang for helpful discussions and assistance with the measurements and data analysis. This work was supported in part by Aerospace Medical Research Laboratory Contract SCEE-HER/86-14.

REFEFENCES

Calvet, J., Calvet, M.C., and Scherrer, J. (1964). Etude stratigraphique de l'activité EEG spontanée. Electroenceph. clin. Neurophysiol. 17, 109-125.

Carelli, P., Foglietti, V., Modena, I., and Romani, G.L. (1983). Magnetic study of the spontaneous brain activity of normal subjects. Il Nuovo Cimento 2D, 538-546.

Chatrian, G.E., Bergamini, L., Dondey, M., Klass, D.W., Lennox-Buchthal, M., and Petersén, I. (1974). A glossary of terms commonly used by clinical electroencephalographers. Electroenceph. clin. Neurophysiol. 37, 538-548.

Creutzfeldt, O., and Houchin, J. (1974). Neuronal basis of EEG-waves. In: Rémond, A., Handbook of Electroencephalography and Clinical Neurophysiology, Vol. 2, Part C, The Electrical Activity from the Neuron to the EEG and EMG, The Neuronal Generation of the EEG, Elsevier, Amsterdam, pp. 5 - 55.

Frost, I.D. (1968). EEG-intracellular potential relationships in isolated cerebral cortex. Electroenceph. clin. Neurophysiol. 24, 434-443.

Hämäläinen, M.S., and Ilmoniemi, R.J. (1984). Interpreting measured magnetic fields of the brain: estimates of current distributions. Report TKK-F-A559, Helsinki University of Technology.

Lopes da Silva, F.H., and Storm van Leeuwen, W. (1978). The cortical alpha rhythm in dog: the depth and surface profile and phase. In: Brazier, M.A.B., and Petsche, H., Architecture of Cerebral Cortex, Raven Press, New York, pp. 319 - 333.

Vvedensky, V.L., Ilmoniemi, R.J., and Kajola, M.S. (1986). Study of the alpha rhythm with a 4-channel SQUID magnetometer. Med. & Biol. Eng. & Computing 23, Suppl. Part 2, 11-12.

Williamson, S.J., Pelizzone, M., Okada, Y., Kaufman, L., Crum, D.B., and Marsden, J.R. (1984). Magnetoencephalography with an array of SQUID sensors. In: Collan, H., Berglund, P., and Krusius, M., Eds. Proceedings of the Tenth International Cryogenic Engineering Conference, Butterworth, Guildford, pp. 339-348.

Fragmenting pictures on the Apple Macintosh computer for experimental and clinical applications

Joan Gay Snodgrass

Bradford Smith

and

Kelly Feenan

Department of Psychology

New York University

and

June Corwin

NYU School of Medicine

and

Psychiatry Service

New York VA Medical Center

Address correspondence to:

Joan Gay Snodgrass

6 Washington Place, Room 857

New York, NY 10003

(212)598-2327

Abstract

A set of procedures implemented in Microsoft BASIC is described which creates fragmented versions of pictures scanned into the Apple Macintosh, stores them as resource files, and presents them in a computerized perceptual memory test. A total of 150 pictures were selected from the Snodgrass and Vanderwart (1980) set for fragmentation. The perceptual memory test provides for five forms of 30 pictures each divided into two sets of 15 which serve alternately as the training or old set and the new set. A training set of 15 pictures is presented for identification during the first (training) phase of the test. The second (test) phase presents the training pictures again randomly mixed with 15 new pictures for identification.

The performance of 100 subjects on the memory test is presented, along with results for each form. Overall, subjects show improvement on the task with practice (skill learning), indexed by a decrease in thresholds from the train to the new set. Subjects also show large savings for the repeated pictures (perceptual learning), indexed by a decrease in thresholds from the new to the old set.

Fragmenting pictures on the Apple Macintosh computer for experimental and clinical applications

This paper describes a set of procedures for fragmenting picture stimuli, storing them as resource files, and presenting them in a computerized test of perceptual learning which runs on an Apple Macintosh plus. The test follows the Gollin Picture Test (Gollin, 1960) in spirit. In the Gollin test, subjects are shown a series of pictures from which fragments have been deleted, starting with the most fragmented level. Increasingly more complete versions of each picture are shown until all pictures can be identified. Subjects are then retested on the old items to measure perceptual memory. Although the test was initially developed for use with children, it has been used extensively with clinical populations (e.g., Corkin, 1982), following the demonstration by Warrington and Weiskrantz (1968) that amnesic patients show substantial learning and retention as measured by decreased thresholds for repeated pictures after a delay as long as three months. This preserved learning ability in amnesics, along with demonstrations in normal subjects that effects of many variables are dissociated in perceptual compared with episodic learning tasks, have led several investigators to postulate the existence of two or more independent memory systems (e.g., Graf, Squire, & Mandler, 1984; Jacoby & Dallas,1981; Jacoby & Witherspoon, 1982; Tulving, 1984), although Jacoby (1983) presents an alternative interpretation of such dissociations.

The Gollin set of fragmented pictures consists of 20 pictures of common objects and animals, and the present set consists of 150 pictures of common objects and animals taken from the Snodgrass and Vanderwart (1980, hereafter S & V) set of standardized pictures. Each picture in the Gollin set has five levels of fragmentation, while the present set has eight levels.

Several properties of the Gollin set were incorporated in the present set. First, as in Gollin's stimuli, fragmentation is cumulative, so that all fragments in a more degraded version are present in a less degraded version. Second, the exact same fragments are presented when stimuli are repeated as in the original training. This contrasts with fragmentation algorithms which occur on-line, where randomly selected portions of the picture are deleted each time (e.g., Vokey, Baker, Hayman, & Jacoby, 1986). This also contrasts with any procedure in which stimulus degradation is accomplished through temporal limitations, such as brief flashes of intact stimuli on a tachistoscope or CRT. In these cases, the observer's perceptual system introduces random fluctuations in determining which portions of the degraded stimulus are seen. Finally, a limitation of the Gollin set is that the most fragmented versions are often easily identifiable, leading to ceiling effects. In contrast, in the present set the most fragmented versions are virtually impossible to identify with no previous exposure.

We believe that an important advantage of the present procedure is the ability to control whether the same or different fragments are presented again during memory testing. Comparisons of the two conditions may clarify differences obtained by researchers in the clinical literature with Gollin pictures as opposed to other perceptual tasks.

The remainder of this paper is divided into two parts. In the first, we describe the set of procedures which are used to produce the fragmented pictures, and in the second, we describe the results of a study using the basic task.

Generating the stimuli

The first step was to select candidate pictures from S & V which were considered suitable for fragmenting. We selected pictures having moderate complexity and sufficient area so that distinctly different fragmented images could be created. So, for example, accordion was excluded because of its high complexity and needle was excluded because of its lack of area.

The next step was to input the pictures into the Macintosh. Because the pictures were to be initially saved as MacPaint files, we reduced the size of the original drawings to correspond to the size of a MacPaint window (approximately 246 x 246 pixels). The pictures were digitized using the Thunderscan digitizer, an inexpensive peripheral device for the Macintosh which replaces the ribbon cartridge on the Apple ImageWriter and scans an image on a sheet inserted into the printer carriage. The picture is saved as a MacPaint file. Within MacPaint, the picture can be cleaned up with the aid of Fat Bits, and then is centered within a selection square approximately 246x246 pixels in dimension, and copied to the scrapbook.

Pictures are copied or cut from the scrapbook onto the clipboard, and then read from the clipboard by an MBASIC program using the "OPEN CLIP:PICTURE" statement. The pictures are saved in MBASIC sequential files as string variables consisting of encoded graphics instructions (PICT format), one picture to a file. This storage format conserves space compared to the storage requirements for a bit-mapped image. The average storage requirements for the set of 260 pictures from the S & V set stored in PICT format as MBASIC sequential files is approximately 3K. However, a disadvantage of the PICT format is that it takes longer to draw on the screen than a bit-mapped image.

Subsequent forms of the pictures are saved as resource files, using libraries available from Clear Lake Research. Resource files have the following advantages over either sequential or random access files: (1) they require less storage space than the other types; (2) they permit pictures to be accessed individually by either name or identification number; and (3) they permit replacement of a single picture (Brooks, 1985).

<u>The fragmentation process.</u> In order to produce fragmented pictures that are sufficiently difficult to identify, we found it necessary to delete fairly large blocks of pixels. By a process of trial and error, we chose a 16 x 16 pixel block size, although other sizes in that vicinity would also have been acceptable.

In order to delete cumulatively, and to ensure that each successive fragmentation level has fewer fragments than the next lower level, it was necessary to identify which 16 x 16 blocks contain information. Because each picture is drawn within a 246 x 246 window, there are a total of 256 blocks (blocks in the rightmost column are truncated).

The fragmentation program lays out a grid of 16 x 16 blocks, determines which blocks contain black pixels, and stores the locations of these critical blocks. Then the program randomly selects increasing proportions of critical blocks to be erased according to an exponential function, to produce eight levels of fragmentation per picture. Level 8 is the complete picture, and level 1 is the most fragmented picture. The parameter of the exponential function is adjustable to produce more rapid fragmentation at higher values or slower fragmentation at lower values. It is also possible to vary the number of levels in the fragmentation series.

Because the deleted blocks are determined randomly, a very large set of possible fragmentations is possible. In constructing the fragmented stimuli, we ran through several fragmentations until we identified a "good" one. The sequence of random numbers which determined which blocks were to be deleted at each level were saved so that a particular fragmentation series could be reproduced. Our criteria for a good fragmentation were first, that the general outline of the picture was retained at the most fragmented level, and second, that critical features which would identify a picture (such as the eyes or tail of an animal) were deleted as soon as possible.

Once an acceptable fragmentation series was identified, the random number array was used to create and store eight levels of fragmentation for each picture. The fragmented pictures are stored in resource files in folders on a stimulus disc. The stimulus disc also contains a masterfile for each set, which has the names and identifying numbers (from S & V) for each of the pictures in a set, as well as the pictures themselves. The memory test program uses the masterfile to determine the identity of the pictures to be used in a particular condition. The stimulus disc also contains resource files with alternative correct names (variant files) for each picture. These are used by the program to determine whether the subject's response matches any of the correct names for a particular picture. The experimenter can update the variant files by adding additional names to the correct set.

The program disc contains the program for running the memory test, a customized library of Clear Lake Research statements which access machine language routines, the program for updating variant files, and a program for accessing the stored data from individual subjects. The analysis program lets the user print out data from individual subjects and write the data to the clipboard, from which it can be pasted to a spreadsheet. The data consists of thresholds for each picture within the train, new, and old sets, sorted by identification number (and thus alphabetically by the name of the picture). The mean thresholds for the train, old and new conditions are also computed. Incorrect responses are not stored.

Figure 1 shows three examples of fragmented pictures at selected levels of fragmentation. The levels shown, from top to bottom, are level 8 (complete picture), 6, 4, and 2 (one level above the most fragmented level).

Insert Fig. 1 about here

The memory test. We created five forms of the memory test by dividing the set of 150 pictures into 10 sets of 15 pictures equated on name agreement, rated familiarity, and rated complexity according to the S & V norms, and with approximately the same distribution of exemplars across categories. A particular form of the test uses two sets of 15 pictures, one set for the training or old stimuli, and one set for the new stimuli. The five forms are numbered 1 through 5, and the two sets within each form are labeled a and b. By using a or b as training stimuli (and b or a as new stimuli), a total of 10 forms can be created, although a particular subject can only be exposed to five forms (e.g., 1a *or* 1b) without repeating items across tests.

In the first phase of the memory test, a form and set is selected by the experimenter to serve as old stimuli and these pictures are randomly ordered by the program. Picture identification thresholds are measured by the ascending method of limits. The most fragmented version of each picture is presented and the subject either attempts to identify it (by typing its name on the keyboard) or presses the return key to go on the next level. A name is correct if it matches one of the listed names for that concept stored in the variant files. Incorrect names are not penalized, so there is no penalty for guessing. The experimenter can update a variant file by adding an additional name that he or she wishes to be considered as correct. The variant files were constructed by using the names that occurred more than once in the S & V norms, and then adding those which more than one subject responded with in the normative study reported here.

When the picture has been correctly identified, the program stores the level of fragmentation at which the training item was identified and moves on to the next picture in the random sequence. After all 15 pictures have been correctly identified, subjects can be given a distractor task (we used a 10-minute cancellation of 9's paper and pencil test), or subjects can proceed immediately to the test phase.

In the test phase, subjects are presented with a test set of 30 pictures presented in a random order determined by the program. The test set consists of the 15 pictures used in the training series (the old set) and a set of 15 new pictures (the new set). Subjects go through the same procedure of identifying each picture by typing in its name or pressing the return key to get the next level. As in the training phase, there is no penalty for guessing.

The data of interest are the identification thresholds for the training stimuli, during the first phase, and the identification thresholds for the old and new stimuli during the test phase. Learning the task (skill learning) is indexed by a decrease in thresholds between the training and new stimuli (train minus new). Learning the items (perceptual learning) is indexed by a decrease in thresholds between the new and old items in the test (new minus old). The perceptual learning measure thus excludes improvement in performance due to skill learning.

The memory testing program provides the option of either presenting or not presenting the complete picture in the training phase. This could be implemented to compare the effects of complete stimulus priming on subsequent test identification thresholds. In the normative study reported here, we did not present the complete picture during training (the complete picture is never presented during testing).

Results of the normative study

In order to test the procedure and to obtain normative data on the 150 pictures, we ran the basic memory task with 100 subject volunteers who participated to fulfill a course requirement in introductory psychology. Subjects were randomly assigned to each of the 10 versions of the test (five forms by a and b sets) until 10 subjects had participated in each.

Insert Table 1 about here

Table 1 presents mean thresholds for the train, new, and old stimuli for the a and b versions of each of the five forms. The task is scored so that good performance is accompanied by low numbers. A threshold of 1 means the stimulus was identified at its most fragmented level, and a threshold of 8 means the stimulus was identified as its most complete level. Sets a and b within each form are paired such that one set serves alternately as training and new stimuli. So, for example, 1b serves as the new stimuli for 1a and vice versa.

It is clear that subjects improve with practice on the task (there is an average decrease of 0.203 level from train to new across sets). However, not all subsets show this decrease. Sets 1b, 2a, and 5a all show an increase in threshold between the train and new sets. In two of the three cases, the problem has to do with initial inequality between sets. Set 1a is harder than 1b, and set 5b is harder than 5a, so subjects' skill learning is masked by set differences. .

It is also clear that subjects improve markedly on the repeated items, with an average decrease in threshold from new to old of almost two fragmentation levels. All forms and sets within forms show this decrease.

Several analyses of variance were performed in order to separately investigate the learning component and the forms or item specific effects. To evaluate the learning component with minimal influence of set differences, we combined the a and b forms of each test (to eliminate differences between set a and set b) and performed a 3 (training) x 5 (form) mixed analysis of variance. There was a significant effect of form [$F(4, 95) = 3.55$, $p < .01$], a highly significant effect of training [$F(2, 190) = 142.65$, $p < .001$], and no interaction between form and training [$F(8, 190) = 1.31$]. The results of Tukey HSD post hoc comparisons showed that both memory components were significant at the .01 level. Subjects showed significant skill learning because new thresholds were significantly lower than train thresholds. And subjects showed significant perceptual learning because old thresholds were significantly lower than new thresholds. The only significant form difference was that between Form 1 and Form 3. The lack of a significant interaction is not particularly surprising given that across forms, item effects are counterbalanced between the train and new sets, and thus average out when combined. Figure 2 presents the average thresholds as a function of training for each form.

Insert Figure 2 about here

To investigate differences between a and b sets within forms in more detail, we carried out individual 3 (training) by 2 (set: a vs b) analyses of variance on each of the five forms. An ideal form would show no significant effect of set, a significant effect of training, and no interaction of set with training. All forms show a significant effect of training [all $F(2, 36)$'s > 215]. Only two forms (3 and 4) show no effect of set and no interaction. Forms 1 and 5 both show an insignificant effect of set but a highly ($p < .0001$) significant interaction, while Form 2 shows both a significant effect of set ($p = .006$) and a significant interaction ($p=.0013$).

In order to optimize the use of this memory test, particularly for use in clinical trials (where several equivalent forms of a test are required to evaluate treatment effects), it will be necessary to adjust items within forms 1, 2, and 5 so that they are more nearly equated for difficulty of initial identification in their a and b sets. We are presently engaged in doing just that. In the meantime, however, forms 3 and 4 may be used with a fair amount of confidence that items within their a and b sets are equivalent. This provides for two independent forms for within subjects designs and four quasi-independent forms for between subjects designs (i.e., 3a, 3b, 4a, and 4b).

Discussion

The experimental task presented here makes it possible to separately measure two components of memory: skill learning, measured by an improvement on the task across different items, and perceptual learning, measured by an improvement on the task across the same items. Skill learning has also been called procedural memory in the literature, and perceptual learning has also been called implicit memory or perceptual fluency. Although the skill learning component is significant in this task, perceptual learning is almost an order of magnitude larger. However, we do not believe that the perceptual learning component as measured in this task is independent of episodic memory influence. Some of the perceptual learning effect could be due to explicit episodic memory for the fragments themselves (because the same fragments are presented on the old trials as were presented on the train trials), or to memory for the names of the training items which occur as lucky guesses during test. We are presently investigating both of these possibilities.

Availability of the memory test

A disc (800K) containing 250 of the 260 S & V pictures as MBASIC files is available from the first author for $5.00.

The complete perceptual memory test is available on six 800K discs. One disc contains the programs for running the experiment and analyzing the data and the remaining five discs contains the five memory test forms. These are available from the first author for $30.00. The user must have MBASIC (version 2.1 or higher) to run these programs.

In addition, Brooks (1985) provides all 260 S & V pictures on two discs as MacPaint files.

References

Brooks, J. O., III (1985). Pictorial stimuli for the Apple Macintosh computer. *Behavior Research Methods, Instruments, & Computers, 17,* 409-410.

Corkin, S. (1982). Some relationships between global amnesias and the memory impairments in Alzheimer's disease. In S. Corkin, et al. (Eds.), *Alzheimer's Disease: A report of Progress* (pp. 149-164). New York: Raven Press.

Gollin, E. S. (1960). Developmental studies of visual recognition of incomplete objects. *Perceptual and Motor Skills, 11,* 289-298.

Graf, P., Squire, L. R., & Mandler, G. (1984). The information that amnesic patients do not forget. *Journal of Experimental Psychology: Learning, Memory, and Cognition,10,* 164-178.

Jacoby, L. L. (1983). Perceptual enhancement: Persistent effects of an experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 21-38.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 3,* 306-340.

Jacoby, L. L., & Witherspoon, D. (1982). Remembering without awareness.*Canadian Journal of Psychology, 36,* 300-324.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for naming agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 174-215.

Warrington, E. K., & Weiskrantz, L. (1968). New method of testing long-term retention with special reference to amnesic patients. *Nature, 217,* 972-974.

Vokey, J. R., Baker, J. G., Hayman, G., & Jacoby, L. L. (1986). Perceptual identification of visually degraded stimuli. *Behavior Research Methods, Instruments, & Computers, 18,* 1-9.

Tulving, E. (1985). How many memory systems are there? *American Psychologist, 4*, 385-398.

Footnote

This research was supported in part by a Challenge Grant from New York University, a University Research Initiative Program grant from the Air Force Office of Scientific Research, and the Veterans Administration. Requests for reprints should be sent to J. G. Snodgrass, Department of Psychology, New York University, 6 Washington Place, New York, NY 10003.

Table 1

Mean threshold values by set and training condition

set	train	new	old	*train-new	**new-old
1a	5.23	4.49	3.25	0.75	1.23
1b	4.77	5.19	2.66	-0.42	2.53
2a	5.02	5.27	2.75	-0.25	2.51
2b	4.87	4.38	2.28	0.49	2.10
3a	4.52	4.33	2.33	0.19	2.00
3b	4.57	4.38	2.46	0.19	1.92
4a	4.88	4.41	2.61	0.47	1.80
4b	4.92	4.78	2.92	0.14	1.86
5a	4.39	5.11	2.56	-0.72	2.55
5b	5.07	3.89	2.81	1.19	1.08
Mean	4.82	4.62	2.66	0.20	1.96

*measure of skill learning

**measure of perceptual learning

Figure 1. Examples of fragmented images at levels 8 (complete), 6, 4, and 2.

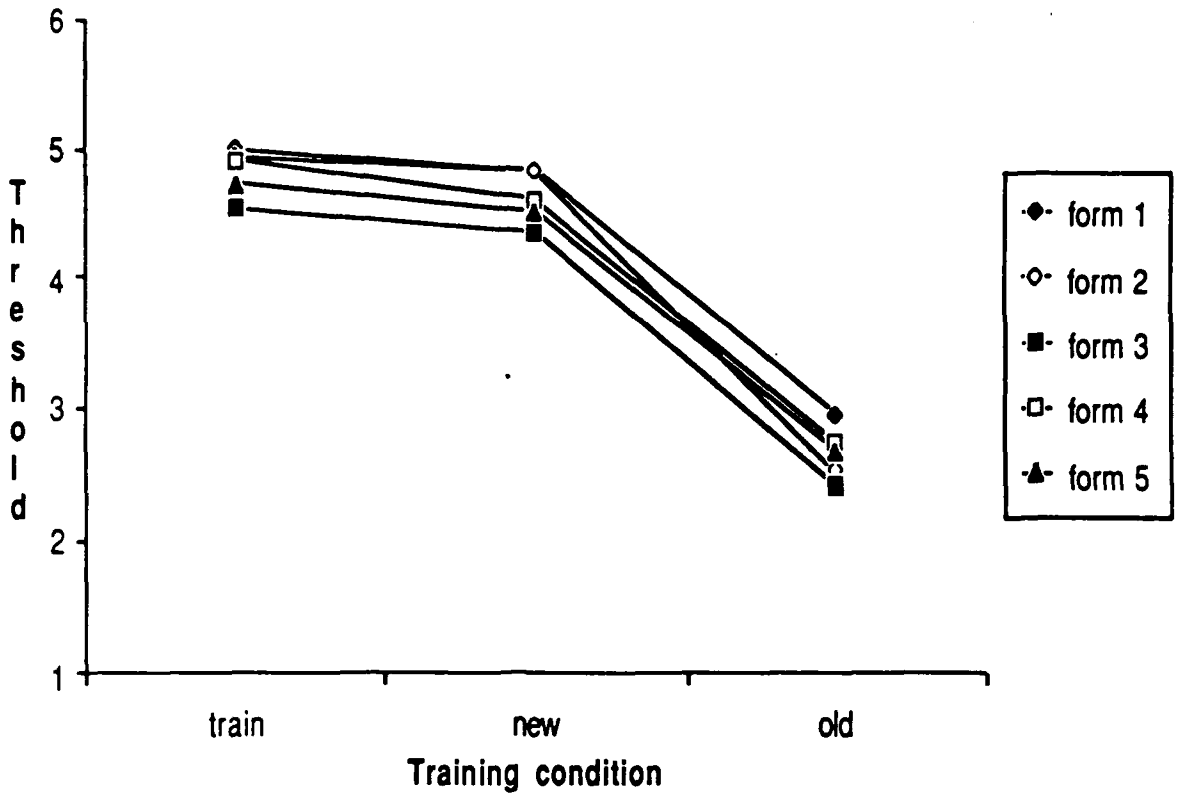


Figure 2. Mean identification thresholds for the three training conditions by form (1=most fragmented and 8 = complete).

AFOSR-TR- 87-1710

Pragmatics of measuring recognition memory: Applications to dementia and amnesia

Joan Gay Snodgrass
New York University
and
June Corwin
NYU School of Medicine and
New York Veterans Administration Medical Center

Direct correspondence to:
Joan Gay Snodgrass
New York University
Department of Psychology
6 Washington Place, Room 857
New York, N.Y. 10003

RUNNING HEAD: Pragmatics of measuring recognition memory

SUMMARY

The present paper has two purposes. The first is to describe four theoretical models of yes-no recognition memory, and present their associated measures of discrimination and response bias. These models are then applied to a set of data from normal subjects to determine which pairs of discrimination and bias indices show independence between discrimination and bias. The following models demonstrated independence: a two-high-threshold model, a signal detection model with normal distributions using d' and C (rather than beta), and a signal detection model with logistic distributions and a bias measure analogous to C. C is defined as the distance of criterion from the intersection of the two underlying distributions.

The second purpose is to use the indices from the acceptable models to characterize recognition memory deficits in dementia and amnesia. Young normal subjects, Alzheimer's disease (AD), and parkinsonian dementia (PD) patients were tested with picture recognition tasks with repeated study-test trials. Huntington's disease (HD), mixed etiology amnesics, and age-matched normals were tested by Butters et al. (1985) using the same paradigm with word stimuli. Demented and amnesic patients produced distinctly different patterns of abnormal memory performance. Both groups of demented patients showed poor discrimination and abnormally liberal response bias for words (HD) and pictures (AD and PD), whereas the amnesic patients showed the worst discrimination but normal response bias for words. Although both SDT and two-high-threshold discrimination parameters showed identical results, the bias measure from two-high-threshold model was more sensitive to change than C from SDT.

Three major points are emphasized. First, any index of recognition memory performance assumes an underlying model. Second, even acceptable models can lead to different conclusions about patterns of learning and forgetting. Third, efforts to characterize and ameliorate abnormal memory should address both discrimination and bias deficits.

A common method for measuring memory performance is by use of a recognition memory test. Here we consider the yes-no form of recognition memory test in which studied (old) items are presented mixed with distractor (new) items and the subject's task is to decide whether each item is old (by responding "yes") or new (by responding "no"). Performance on a yes-no recognition test is summarized by two measures: the hit rate, the probability that the subject classifies an old item as old, and the false alarm rate, the probability that the subject classifies a new item as old. A given model of the recognition memory process specifies how these two measures are to be combined to characterize memory performance.

In 1970, both Banks (1970) and Lockhart and Murdock (1970) published review papers on the use of signal detection theory (SDT) in measurement of human memory. Since then, researchers in the clinical areas have conscientiously adopted the indices derived for SDT to evaluate memory in clinical populations (e.g., Branconnier, Cole, Spera, & DeVitt 1982; Butters, Wolfe, Martone, Granholm, & Cermak, 1985; Mohs & Davis, 1982). In contrast, reseachers investigating normal memory often use a simpler measure of discrimination such as hits minus false alarms or the nonparametric measure A' (Gillund & Shiffrin, 1984; Tulving & Thomson, 1971).

Much of this literature in both normal and abnormal populations focuses on evaluation of discrimination with relatively little attention paid to the other half of the recognition story, response bias. However, in the abnormal literature it is becoming increasingly clear that abnormal bias is an important component of abnormal memory (Branconnier et al., 1982; Butters et al., 1985). In fact, Branconnier et al. demonstrated that false alarm rate fared almost as well as d' in discriminating between Alzheimer's disease (AD) and normal elderly memory.

In this paper, we re-examine the models underlying the use of various indices of recognition memory, and make some suggestions for appropriately measuring both bias and discrimination in normal and abnormal populations. The basic desiderata of such indices are twofold: first, that the discrimination index be invariant across explicit manipulations of bias; and second, that the bias index be invariant across explicit manipulations of discriminability. Otherwise, comparisons between populations having different discrimination performance may on the one hand yield spurious differences in bias, or on the other fail to detect differences in bias which are present. We also note that Murdock (1974, pp. 34-35) proposed and tested a similar criterion for evaluating the signal detection model of recognition memory.

We will first describe four models of the recognition memory process and their associated indices. A critical feature of any model of recognition memory is that it defines a potentially infinite set of hit/false alarm rate pairs which yield equivalent discrimination across all levels of bias. These equal discrimination functions have been called iso-memory functions or MOCs (memory-operating characteristics). We will use the term iso-memory function in the remainder of this paper. In addition, each model defines a set of iso-bias functions, representing sets of hit/false alarm rate pairs which yield equivalent bias across all levels of discrimination.

After describing the iso-memory and iso-bias functions for each model, we next apply these four models to data from an experiment designed to independently manipulate discrimination and bias. The degree to which a model's discrimination and bias indices show the desired invariances will determine the acceptability of the indices (and therefore the model).

Finally, we will use the acceptable indices to characterize learning and forgetting of demented, amnesic, and normal subjects in repeated study-test tasks with picture and word stimuli.

The Recognition Memory Task

All indices of recognition memory performance considered here are based on a single pair of hit and false alarm rates, as would be obtained in a yes-no recognition memory test procedure. In conformity with common usage, we define a hit as a "yes" response to an old item, and the hit rate, H, as the conditional probability of responding "yes" to an old item, P("yes"/old). A false alarm is defined as a "yes" response to a new item, and the false alarm rate, FA, as the conditional probability of responding "yes" to a new item, P("yes"/new). Similarly, a correct rejection is defined as a "no" response to a new item, and the correct rejection rate, CR, as the conditional probability of responding "no" to a new item, P("no"/new). A miss is defined as a "no" response to an old item, and the miss rate, M, as the conditional probability of responding "no" to an old item, P("no"/old). Because the sum of the hit and miss rates is 1.0, the hit rate is sufficient to describe what happened on old item trials . Similarly, because the sum of the false alarm and correct rejection rates is also 1.0, the false alarm rate is sufficient to describe what happened on new item trials. The pair of hit and false alarm rates thus completely summarize the data of a single subject in a single condition of a yes-no recognition memory test.

Correcting data in the stimulus-response matrix

Measures for the two SDT models are undefined for hit rates of 1.0 or false alarm rates of 0 because the corresponding z-scores are infinite. Accordingly, we have corrected all hit and false alarm rates by adding 0.5 to each frequency and dividing by N +1, where N is the number of old or new trials. This correction is recommended for log-linear models (e.g., Upton, 1978). For consistency, we recommend applying the correction routinely, even in the absence of 0's and 1's, and even when SDT measures are not calculated. Figure 1 presents the stimulus-response matrix with the correction shown. This transformation has been applied to all the data reported here.

Insert Fig. 1 about here

The Four Models of Recognition Memory

Signal Detection Theory (SDT)

The signal detection model for recognition memory proposes that items presented for test in a recognition memory task lie along a continuum of familiarity or memory strength. Some old items will have high familiarity, most will have medium familiarity, and some will have low familiarity. New items (distractors) will have similarly distributed levels of familiarity but with a lower mean level than the old items. Theoretically, the distributions of familiarity of old and new items will always overlap. The distribution of familiarity for old items corresponds to the signal + noise distribution of sensory SDT, and the distribution of familiarity for new items corresponds to the noise alone distribution of sensory SDT.

It is assumed that subjects cannot directly determine whether or not an item is old. Rather, the subject is aware only of an item's familiarity or strength value. Thus, the subject must set some criterion value of familiarity, called x_c, such that items whose familiarity exceeds the criterion are called "old" and those that fail to exceed the criterion are called "new." The proportion of old items exceeding the criterion is estimated by the hit rate, and the proportion of new items exceeding the criterion is estimated by the false alarm rate. The criterion is assumed to be under the subject's control, and thus an important feature of sensory SDT experiments is explicit manipulation of the criterion via payoff matrices or presentation probabilities of signal and noise trials. A payoff matrix defines the costs for errors (false alarms and misses) and rewards for correct responses (hits and correct rejections).

SDT with Normal Distributions

In the classic version of SDT, both distributions are assumed to be unit normal curves. In an alternate version described in more detail below, the distributions are assumed to be logistic in form. The following indices are based on normal distributions. The corresponding indices for logistic distributions are given later.

Discrimination index (d'). According to this model, a subject's ability to discriminate between old and new items is given by d', the distance between the means of the old and new distributions in units of the common standard deviation. We assume throughout that the standard deviation of the old distribution, σ_o, equals the standard deviation of the new distribution, σ_n, and denote this common standard deviation as σ. This theoretical definition of d' can be expressed as:

$$d' = (\mu_o - \mu_n)/\sigma$$

where μ_o is the mean of the old distribution and μ_n is the mean of the new distribution.

Given a hit rate <1 and a false alarm rate >0,and the unit normal assumption, d' can be estimated by the z-score of the false alarm rate minus the z-score of the hit rate, or:

(1) $$d' = z_{FA} - z_H$$

where z_H is the z-score in the old distribution having H proportion above it and z_{FA} is the z-score in the new distribution having FA proportion above it. For moderate biases (H >.5 and FA < .5), z_{FA} will be positive and z_H will be negative, so that equation (1) has the effect of adding their absolute values.

Bias indices (ß and C). This model generates several plausible bias measures, each with its own theoretical rationale. The purpose of any bias measure in this model is to locate the criterion, x_c, which is a dimensionless quantity. We consider two bias measures for the SDT model. The likelihood ratio measure, ß, locates the criterion by the ratio of the heights of the old and new distributions, while the intersection measure, C, locates the criterion by its distance from the intersection of the two distributions, as shown in Figure 2.

Insert Figure 2 about here

The first bias measure, ß (beta), is the ratio of the density of the old distribution at the criterion divided by the same density for the new distribution. It is calculated as

$$\beta = f_o(z_H)/f_n(z_{FA})$$

where f_o is the height of the normal distribution over z_H and f_n is the height of the normal distribution over z_{FA}.

This bias index is known as the likelihood ratio because it is the ratio of the likelihood of obtaining an observation equal to the criterion given an old item to the likelihood of obtaining this observation given a new item. A subject with a perfectly neutral response bias will set his criterion at the intersection of the old and new distributions, yielding a ß of 1. More liberal criteria lie to the left of the intersection point, yielding ß's <1 while more conservative criteria lie to the right of the intersection point, yielding ß's >1. Because symmetric placements of the criterion around the intersection point produce reciprocal values of ß, ß must be transformed before analysis to produce interval-scale data. The usual transformation is to take the natural logarithm, as shown below:

(2) $$\ln(\beta) = \ln[f_o(z_H)/f_n(z_{FA})]$$

Thus, ln ß is 0 for a neutral criterion, negative for liberal criteria, and positive for conservative criteria.

Green and Swets (1974) show that ß is appropriate if a subject wishes to maximize expected winnings under a given payoff matrix. However, Lockhart and Murdock(1970) point out several problems with this measure when heterogenously memorable stimuli are used in an experiment. We consider their reservations in more detail below.

An alternative way of locating x_c is to define its location relative to some zero-point along the familiarity axis. One possible zero point is the mean of the new distribution. This locates the criterion within the new distribution, without regard to its location within the old distribution, and as such is determined only by the FA rate. Banks (1970) has termed this criterion index C_j, and points out that it is only valid for hit/false alarm pairs which lie on the same iso-memory function (that is, points indicating equal discrimination).

C_j can easily be calculated as z_{FA}, the z-score having probability FA above it in a normal distribution. However, because we will be concerned with bias measured across different levels of discriminability, we do not consider this measure further, except to note that it has been used in the abnormal memory literature (e. g., Mohs & Davis, 1982).

A more attractive measure of bias, which we call C, locates the criterion relative to the intersection of the old and new distributions. The intersection point defines the zero-point, and distance from the criterion is measured in z-score units. A completely neutral bias has a C value of 0, conservative biases produce positive C values, and liberal biases produce negative C values. (Note that this is exactly opposite to ln ß where conservative biases produces negative values of ln ß and liberal biases produce positive values of ln ß.)

To compute C, we first determine the distance of the criterion from the mean of the new distribution, which is set equal to 0. This is given by z_{FA}, the z-score having probability above it equal to the false alarm rate. To change the 0 point from the mean of the new distribution to the intersection of the old and new distributions, we subtract $d'/2$ from z_{FA}. The formula is shown below:

$$C = z_{FA} - d'/2 = 0.5(z_{FA} + z_H) \quad (3)$$

C and ß compared. Lockhart and Murdock (1970) point out several problems with the use of ß as a criterion measure. When heterogenously-memorable stimuli (such as high and low imagery words) are presented for test, investigators usually assume that the criterion is set with respect to the appropriate old *and* new familiarity distributions -- i.e., the familiarity of a high imagery word is compared to the distribution of both old and new high imagery words. This assumption, however, requires that the subject first classify the item as a member of the appropriate class, and then somehow locate its relevant new distribution, before making such a decision. Yet locating the relevant new distribution would appear to require prior perfect knowledge of whether the item is old or new, the very decision we are asking the subject to make as the test of learning.

In contrast, the use of C requires only that the subject know the familiarity value of the test item. If we assume, with Glanzer and Adams (1985), that old and new distributions mirror one another in their locations along the familiarity axis, so that new distributions of highly memorable classes of items are lower on the familiarity continuum than new distributions of less memorable classes, this provides a mechanism whereby subjects can use the familiarity value itself as the basis for making the decision. If the increment in oldness is equal to the increment in newness, then the C index of criterion, defined as it is with respect to the intersection of the corresponding old-new distributions, appears to reflect a psychologically plausible mechanism.

SDT with Logistic Distributions

Noreen (1977) has shown that when logistic distributions replace normal distributions in a signal detection model, discrimination and bias indices are identical to those from two other important models of choice behavior -- Luce's (1959, 1963) choice model and Link and Heath's (1975) random walk model. Additionally, indices based on signal detection theory with logistic distributions are considerably easier to calculate than d', ß, and C, thereby reducing the computational labor and chance for error. For these two reasons, we present these alternative discrimination and bias indices here. Later we show that the two sets of SDT-based indices yield equivalent results when applied to a variety of data sets.

The logistic distribution function (the cumulative of the density function) in standard form ($\mu = 0$, $s = \pi/\sqrt{3}$) is:

$$F(x) = (1 + e^{-x})^{-1} = e^{x} / (e^{x} + 1)$$

where x is a standard score similar to the z-score of the normal distribution.

The beauty of the logistic distribution is the simplicity with which the distribution function can be related to x (this is, of course, not a property shared by the normal distribution). By taking the logit, or log-odds transformation, of F(x), we are left with x itself, as shown below:

$$\ln [F(x)/(1-F(x))] = \ln \{[e^{x} /(e^{x} + 1)]/[1/(e^{x} + 1)]\}$$

$$= \ln e^{x} = x$$

The recovered score, x ,is a standard score comparable to z for the normal distribution.

To recapitulate, in the logistic distribution, the (natural) log of the ratio of the area above an x-score to the area below the x-score produces the score itself. So, for example, the log of the hit rate divided by its complement the miss rate produces the value of x corresponding to the criterion, x_c, within the old item distribution. The corresponding operation on the false alarm rate and its complement (the correct rejection rate) produces the value of x_c within the new distribution. This property of the logistic distribution is used to derive the following measures (details of the derivations can be found in Noreen,1977):

Discrimination index (logistic). A d' -like measure, d_L, is computed as follows:

(4) $$d_L = \ln \{[H(1\text{-}FA)]/[(1\text{-}H)FA]\}$$

Bias indices (logistic). A ß-like measure can be computed as follows (the natural logarithm is computed for the same reasons as for the SDT normal model):

(5) $$\ln(\beta_L) = \ln\{ [H(1\text{-}H)]/[FA(1\text{-}FA)]\}$$

A C-like measure can be computed as follows:

(6) $$C_L = 0.5 [\ln\{[(1\text{-}FA)(1\text{-}H)]/[(H)(FA)]\}]$$

Threshold Theories of Recognition Memory

Threshold models, in contrast to signal detection models, do not assume a continuum of memory strengths, but rather define discrete memory states. In one-high threshold theory, one threshold defines two memory states -- recognition and nonrecognition. If an old item crosses the subject's memory threshold, it will be correctly identified as old. If an old item fails to exceed the threshold, it may be identified as old or new depending upon the subject's response bias. New items can never cross the threshold in this model; this is why the model is termed "high" threshold. New items can only be misidentified as old by guessing on the basis of response bias from the nonrecognition state.

In a two-high threshold model, there are two memory thresholds, one for old items and one for new items. Two thresholds define three possible memory states - old recognition, new recognition and uncertainty. Old items crossing the old recognition threshold will always be identified as old and new items crossing the new item threshold will always be identified as new. As in the one-high threshold model, new items can never cross the old item threshold and old items can never cross the new item threshold. Items in the uncertain state will be classified as old or new depending on the subject's response bias. Thus false alarms and misses always occur from the uncertain state. Because the one-high threshold model is easily falsified by data, we consider here only the two-high threshold model.

Two-high Threshold Model

Discrimination index (P_r) There are two discrimination indices in this model: P_o, the probability that an old item will exceed the old recognition threshold, and P_n, the probability that a new item will exceed the new recognition threshold. From a single hit and false alarm rate, it is impossible to obtain separate estimates for the two thresholds. Accordingly, we make the simplest possible operational assumption -- that the two thresholds are equal -- and define this common threshold as P_r.

The mirror effect in recognition, documented by Glanzer and Adams (1985), suggests that this equivalence assumption may be warranted. The mirror effect is the finding that as hit rates increase across various manipulations, the corresponding false alarm rates decrease in a "mirror" or inverse fashion.

The hit rate is composed of a certain proportion of true recognitions plus lucky guesses from the uncertain state. Because false alarms are only generated from the uncertain state, the false alarm rate is a direct estimate of the probability of saying "Yes" when uncertain. Thus, the hit rate is related to P_r by:

$$H = P_r + FA$$

By rearranging terms, we get an estimate of P_r as:

(7) $$P_r = H - FA$$

Bias index (B_r). This bias index is the probability of saying "yes" to an item when in the uncertain state. False alarms occur when two things happen: the subject fails to recognize the item as new (which occurs with probability $1-P_r$), and the subject guesses "yes" when uncertain (which occurs with probability B_r). Thus the false alarm rate is related to B_r by:

$$FA = (1-P_r)B_r$$

Rearranging terms, and replacing P_r, the discrimination index, by H-FA , its estimate, gives us the following expression for B_r expressed in terms of the hit and false alarm rates:

(8) $$B_r = FA/[1-(H-FA)]$$

A value of B_r equal to 0.5 indicates neutral bias, a value greater than 0.5 indicates liberal bias, and a value less than 0.5 indicates conservative bias.

It should be noted that the discrimination index from the two-high-threshold model is used quite frequently in recognition memory studies (e.g., Gillund & Shiffrin, 1984; Tulving & Thomson, 1971) although usually without reference to its underlying model. Indeed, as early as 1938, Woodworth suggested this particular "correction for guessing". Egan (1958) was apparently the first to state the model underlying this particular measure explicitly .

Distribution-free (Nonparametric) Model

Pollack and Norman (1964) proposed as an alternative measure of discrimination the area under an "average" isomemory curve drawn through a single pair of hit and false alarm rates. Use of the area under an isomemory curve follows from the demonstration by Green and Moses (1966) that the area under an iso-memory curve is a good estimate of forced-choice memory performance. Figure 3 presents the areas used in deriving these nonparametric indices, and the computing formulas are from Grier (1971).

Insert Fig. 3 about here

Discrimination index (A'). The rationale behind the computation of A' is that any reasonable iso-memory function through the single hit/false alarm point shown in the unit square must pass through the areas A1 and A2. Accordingly, A' is defined as area B (all of which must lie below any reasonable iso-memory function) plus 1/2 the sum of A1 and A2, or:

$$A' = B + (A1+A2)/2$$

It is computed from a pair of hit-false alarm rates as follows:

for H > FA:

(9a) $A' = 0.5 + [(H-FA)(1+H-FA)]/[(4H(1-FA)]$

When FA > H, the point lies below the chance diagonal, and the formula must be revised as follows:

for FA > H:

(9b) $A' = 0.5 - [(FA-H)(1+FA-H)]/[4FA(1-H)]$

When H = FA, the point lies on the chance diagonal, and A' = 0.5.

Bias indices (B" and B'_H). Several logical bias indices may be constructed. Grier (1971) proposed an index called B" which is the difference in the two areas A1 and A2, divided by their sum. Notice that if A1 < A2, the subject has a bias toward "yes" and if A2 < A1, he has a bias toward "no." As for A', there are two computing formulas: one for the case of H > FA and one for the case of FA > H.

For H > FA:

(10a) $B'' = [H(1-H)-FA(1-FA)]/[H(1-H)+FA(1-FA)]$

When FA > H, the H and FA values are exchanged as follows:

(10b) $B'' = [FA(1-FA)-H(1-H)]/[FA(1-FA)+H(1-H)]$

A second measure, proposed by Hodos (1970) is the difference between the two areas divided by A1, computed by

(11a) $B'_H = 1- \{[FA(1-FA)]/[H(1-H)]\}$ when H < (1-FA) so that bias is conservative and

(11b) $B'_H = \{[H(1-H)]/[FA(1-FA)]\} -1$ when H > (1-FA) so that bias is liberal

When H = FA, $B'_H = 0$

For both measures, a zero value indicates a neutral criterion, a positive value indicates a liberal criterion, and a negative value indicates a conservative criterion. Both bias measures lie between -1 and +1.

Table 1 summarizes computing formulas for the discrimination and bias measures discussed.

Insert Table 1 about here

Comparison of Models

Because three of the four basic models of recognition memory have two alternative bias measures, there are seven models defined by a discrimination-bias index pair. Each of these seven models defines a unique set of iso-memory and iso-bias curves. Table 2 gives the formulas for the iso-memory and iso-bias curves for four of the seven models, and Figure 4 shows iso-memory and iso-bias functions for each model at two levels of discrimination and three levels of bias. The four models are SDT with logistic distributions and the bias measure β_L, SDT with logistic distributions and the bias measure C_L, two-high threshold theory, and nonparametric theory with the bias measure B". Unlike the SDT logistic model, the SDT normal model does not yield closed forms for the iso-memory and iso-bias functions. Therefore, we have omitted SDT with normal distributions, which has the same shape as SDT with logistic distributions. We have also omitted nonparametric theory with B'_H, which has the same shape as that with B".

Insert Table 2 & Fig 4 about here

The iso-memory and iso-bias functions differ in shape, and therefore define different pairs of hit and false alarm rates which yield equivalent discrimination and equivalent bias. With the exception of two-high threshold theory, which predicts linear iso-memory functions, all other models predict curved iso-memory functions. At high levels of discrimination, the nonparametric iso-memory functions asymptote to the hit and false alarm axes quickly. This "flattening" means that there are minimum values of hit rates and maximum values of false alarm rates which limit the attainment of any given value of A'. For example, it is impossible to achieve an A' of .9 with a hit rate less than .6 or a false alarm rate great than .4 (see Figure 4D).

Perhaps the most notable differences among the models occur for the iso-bias contours. Only the iso-bias contours for two-high threshold theory and for SDT with the intersection measure C maintain separation as discrimination decreases (that is, as the iso-bias contours approach the chance diagonal.) Thus, with these models it is possible to observe bias differences among subjects even when they are operating close to chance. The other models (i.e., SDT with the likelihood ratio β and the nonparametric model with either bias index) predict iso-bias contours which converge toward common origins with chance performance. What this means in practice is that bias differences are harder to measure as performance decreases. In the limit, subjects performing at chance who show a maximum yea-saying bias (i.e., 100% hits and false alarms) or a maximum nay-saying bias (i.e., 0% hits and false alarms) will not be distinguishable from subjects showing intermediate levels of bias (including those showing no bias).

To investigate the theoretical relationship between discrimination and bias for each model further, we varied discrimination but kept bias at its maximum possible values (both liberal and conservative) for each value of discrimination. When discrimination is maximum (that is, 100% hits and 0% false alarms), there can be no bias regardless of the model. However, when discrimination is at chance (either 100% hits and false alarms, for a maximum liberal bias, or 0% hits and false alarms, for a maximum conservative bias) we should, in principle, be able to observe this with a reasonable bias index. Yet as shown in Figure 5, two models (SDT with ln ß and the nonparametric model with B") show marked dependence of bias on discrimination at low levels of discrimination. That is, these maxima decrease as discrimination decreases. In the limit, as discrimination decreases to chance, each bias index approaches its neutral value. This demonstration supports our observations about the nonindependence of bias and discrimination in the iso-bias contours.

INSERT FIG. 5 ABOUT HERE

On the basis of these results alone, we are tempted to recommend either SDT with the intersection measure C or the two-high threshold model for use in recognition memory research, particularly in studies in which discrimination varies as a function of population or as a function of experimental manipulation. It has sometimes been assumed that the distribution-free model is preferable on the grounds that it does not make assumptions about the form and nature of the process underlying generation of hits and false alarms. However, because it determines a unique iso-memory and iso-bias function for any possible H, FA pair, we believe it constitutes as strong a model of recognition memory as any other theoretical approach. Furthermore, as we have shown, it shows a marked dependence of bias on discrimination, and therefore violates our independence criterion.

In summary, at a theoretical level, three of the seven models -- the two-high threshold model and both SDT models with the intersection bias measure -- fulfill our criterion of independence between discrimination and bias. In the first experiment, we address the pragmatics of whether theoretical lack of independence produces empirical problems. That is, over the range of levels of performance found in normal subjects across variations in stimulus materials, do the different models mirror their theoretical behavior in actual data?

Experiment 1

In Experiment 1, discrimination was manipulated in a recognition memory task by using high and low imagery word stimuli, and bias was manipulated by using three payoff matrices designed to produce liberal, neutral, or conservative response biases.

Method

Subjects and design. Ten subjects, of whom three were females, volunteered to participate in the experiment as part of an introductory psychology course requirement. The experiment had a 2 (imagery level: low vs high) by 3 (payoff matrix: liberal, neutral, and conservative) factorial within-subjects design.

Materials. A total of 360 words was selected from the imagery norms published by Paivio, Yuille, and Madigan (1968) so as to be equally divided between high and low imagery levels. The high imagery words were selected from those having scale values of 6 or above on a 7-point scale, and the low imagery words were selected from those having scale values of 4 or below. All of the words were of A or AA frequency in the Thorndike and Lorge (1944) frequency counts.

The 360 words were divided into three sets of 120 each, of which half served as old items and half served as new items, with an equal number of high and low imagery items in the old and new sets. Words were typed in upper case letters on 3" x 5" index cards.

Procedure. Each subject participated in three sessions on three successive days, each day under a given payoff matrix. The payoff matrices were presented to the subjects on index cards at the beginning of each test phase, and remained in view throughout. The neutral payoff matrix rewarded each hit and correct rejection by one point and penalized each false alarm and miss by one point. The liberal matrix rewarded hits more than correct rejections (+10 versus +1), and penalized misses more than false alarms (-10 versus -1). The conservative matrix did just the opposite, rewarding correct rejections more than hits (+10 versus +1) and penalizing false alarms more than misses (-10 versus -1). Subjects were motivated to earn a maximum total number of points by the offer of a monetary prize for the best performance.

Each subject participated in three sessions on three successive days, one day under each of the payoff matrices. The test procedure was identical on each day . Subjects were first presented with one set of 60 study items (30 low and 30 high imagery) and given 2 minutes to go through the pack at their own pace. This study period was followed by a 2 minute filled delay. Subsequently, subjects were presented with the day's payoff matrix, which was explained by the experimenter. The subject was then presented with a shuffled deck of old and new items and asked to sort the items into appropriate piles. The test phase was untimed. At the end of each day's session, test performance was calculated and reported to the subject. The orders of word sets and payoff matrices were counterbalanced across subjects.

Results and Discussion

Figure 6 presents hit/false alarm rate pairs for each of the six conditions pooled across subjects. Inspection of this figure suggests that the imagery and payoff manipulations had the intended effects: high imagery words were discriminated better than low imagery words, while the liberal payoff matrix produced more liberal responding and the conservative payoff matrix more conservative responding than the neutral matrix. It appears by inspection that the iso-memory and iso-bias functions on which these data points must lie are most closely approximated by either SDT model with the C intersection bias parameter or by the two-high threshold model because the bias manipulation yields data points which lie on functions converging to the chance diagonal rather than to the origins. We present statistical evidence for this observation below.

Insert Figure 6 about here

Comparison of models

For each model, all possible pairs of discrimination and bias indices were calculated and submitted to 2 x 3 within-subjects analyses of variance with imagery and payoff matrix as the two independent variables and index as the dependent variable. An ideal pair of indices would show the following properties: (1) for the discrimination index, a significant effect of imagery, no effect of payoff, and no interaction and (2) for the bias index: an effect of payoff, no effect of imagery, and no interaction. The assumption here is that because trials are not blocked by imagery, bias remains constant within a session.

Insert Tables 3 & 4 about here

Tables 3 and 4 present mean values of discrimination and bias indices for each of the models. The measures are grouped by the variable across which they are expected to remain constant. Thus, discrimination measures for the high imagery condition for all three payoff matrices are to the left in Table 3, and discrimination measures for the low imagery condition for all three payoff matrices are to the right. To the left of each measure we indicate the degree to which it fulfills the conditions. Two tildes represent ideal performance for the measure.

Considering first the discrimination indices (Table 3), neither hits nor false alarms alone pass the test, but all other discrimination indices appear adequate in showing a main effect of the imagery variable, no effect of payoff, and no interaction. Thus, behavior of the discrimination measures does not permit us to reject any of the "reasonable" models (neither hits alone nor false alarms alone correspond to discrimination measures for any reasonable model).

The bias indices in Table 4 are grouped by payoff matrix. The two bias measures to the left are for the conservative matrix, those in the middle are for the neutral matrix, and those on the right are for the liberal matrix. Here several of the bias indices fail the independence test by showing a significant interaction. In particular, ln ß from both SDT models fails, as do both distribution free measures. The nature of this interaction is just what we would expect on the basis of the theoretical relationships -- namely, that bias is more extreme for high imagery words than for low imagery words even though subjects made their Yes/No decisions to random presentations of the two types of words.

The two models which pass the independence test are two-high threshold and SDT (normal and logistic) using the intersection bias index C. Thus on the basis of these results we would recommend the use of either two-high threshold or SDT measures with the stipulation that C be used in place of ln ß. This deviates from common practice because most investigators measure response bias with ß. These empirical results confirm our previous observations on the theoretical nonindependence of d' and ln ß and of A' and either B" or B'_H.

Experiment 2

In the second experiment, we applied the preferred measures of discrimination and bias to yes-no recognition memory for meaningful pictures in young normal and moderately to moderately severely demented subjects. Both groups received repeated study-test trials with a final delayed test trial. We examined the fate of the models in this paradigm, and explored the differences between normal and abnormal picture recognition memory.

In addition to our previous concerns, we were interested in recognition paradigms for practical reasons. The recent intensive efforts to pharmacologically ameliorate age-related cognitive decline (viz: Bartus, Dean, Beer, & Lippa, 1982) suffer from a lack of appropriate tasks for more severely impaired subjects (Semple, Smith, & Swash, 1982). Additionally, Brinkman and Gershon (1983), in a review of methods for measuring cholinergic drug effects on memory, suggest that visual recognition tasks may be particularly useful in this regard.

Method

Subjects and design. The young normals were 101 undergraduates who volunteered to participate in the experiment as part of an introductory psychology course requirement. Approximately 20 subjects were run with each form of the test.

The demented subjects were 11 patients (one female) at the New York VA Medical Center diagnosed as having either Alzheimer's disease (AD, N=9) or parkinsonian dementia (PD, N=2). Diagnosis was made according to DSM III (American Psychiatric Association, 1980) and NINCDS-ARDA (McKhann, Drachman, Folstein, & Katzman, 1984) criteria for primary degenerative dementia (AD) or dementia with Parkinson's disease (PD). All subjects had moderate to moderately severe memory and cognitive dysfuction with Global Deterioration Scores (GDS of 3 (early confusional stage) to 6 (middle dementia stage) as evaluated by the Guild Memory Test (Crook, Gilbert, & Ferris, 1980) and the Brief Cognitive Rating Scale (Reisberg, Ferris, & Crook, 1982). Of the 11 subjects, one was classified as GDS 3, three as GDS 4, five as GDS 5 and two as GDS 6. Subjects ranged in age from 55 to 79. No subject was receiving treatment with concommitant anticholinergic medication.

Subjects were tested as part of approved treatment protocols for novel therapeutic agents for dementia. None of these investigational new drugs yielded positive therapeutic effects. Subjects were tested repeatedly on from two to five different equivalent forms of the picture memory test, for a total of 119 administrations.

The design for the young normals was a 5(form) by 3 (test trial) mixed design with form between subjects and test trial within subjects. The design for the demented patients was a 5(form) by 4 (test trial) mixed design with the same designation. In addition, the data from the demented and normal subjects were analyzed together (by omitting trial 3 for the demented subjects and collapsing across form) to produce a 2 (subject group) by 3 (test trial) mixed design.

Materials. The stimuli consisted of 250 pictures selected from the Snodgrass and Vanderwart (1980) set of 260 meaningful picture stimuli. These stimuli consist of black on white line drawings of real objects (e.g., asparagus, sailboat). Five forms of the test were constructed for the patient subjects. Each form contained a set of 10 target items, and four sets of 10 distractor items. Items were chosen such that each form and set of stimuli were balanced for familiarity, visual complexity and category membership according to the Snodgrass and Vanderwart (1980) and Battig and Montague (1969) norms. To construct the set used with normal subjects, the 50 items from each form for the patients were concatenated to produce five sets of 50 target ("old") items. The 50 target items were paired with three sets of 50 distractor ("new") items so that each normal subject saw 200 of the 250 stimuli.

Procedure. All subjects were run in a repeated study-test procedure followed by a delayed test trial. The procedure for the patients was modified for the normal subjects to avoid ceiling effects for the latter group. A study phase consisted of presenting the set of n study stimuli (where n = 10 for patients and 50 for normals) presented via cards (patients) or slide projector (normals) for 5 sec each accompanied by the picture's name (patients) or for 1 sec each with no name (normals).

The test phase consisted of presenting the n study items as targets interspersed with an equal number of distractors and asking subjects, for each item, to indicate whether the item was old (by responding "yes") or new (by responding "no"). Test stimuli were presented on cards (for patients) or slide projector (for normals) and responses were oral (for patients) or written (for normals). For both groups, the order of the target stimuli in the test sequence was randomly different from the order on study trial 1, subject to the constraint that no more than three old or new stimuli could appear successively.

A second study-test sequence followed immediately for both groups of subjects. For the patients, the order of the targets remained the same as for study trial 1, while for the normals, the order of the targets was re-randomized. Patients had a third study-test trial, while normals did not. Distractors for each test trial were always new for both groups.

Both groups were then given a filled half-hour delay which for the patients consisted of performing other memory tasks, and for the normals consisted of completing a standard test of spatial relations. At the end of the filled delay, both groups were given a surprise recognition test, in which the *n* targets were presented again mixed with an equal number of new distractors. At the end of the recognition test, normal subjects were additionally asked to recall as many of the targets as they could. These recall data will not be presented here.

Results and Discussion

Effect of form.

To analyze the effect of form for the patients, only those subjects receiving all five forms were used (7 of the 11), so as to avoid effects of heterogeneity of performance. A 4 (trials) by 5 (forms) within subjects analysis of variance was performed separately on the hit and false alarm rates. Neither form nor the form x trials interaction was significant for either analysis. For hits, the main effect of form yielded $F(4, 24) = 2.35$ and for the form x trial interaction, $F(12, 72) = 1.21$. For false alarms, the main effect of form yielded $F < 1$, and the form x trial interaction yielded $F(12,72) = 1.76$. Therefore, all subsequent analyses for patient subjects will ignore form as a variable.

Each normal subject received only one form, so a 4 (trials) by 5(forms) mixed analysis of variance was performed separately on the hit and false alarm rates. Neither measure showed a significant main effect of form or a significant form x trial interaction. For hits, the main effect of form yielded an $F(4, 96) = 1.15$, and for the interaction of form with trials, $F(8, 192) = 1.66$. For false alarms, the main effect of form had an $F(4, 96) = 1.80$, and for the interaction of form with trials, $F < 1$. Therefore, all subsequent analyses for normal subjects will ignore form as a variable.

Learning and Forgetting Across Trials

Recognition performance in terms of preferred discrimination and bias indices will first be analyzed for each group separately, because of the different numbers of learning trials in the two groups, and then the two groups will be combined.

Normals' performance. Table 5 shows mean values of discrimination indices for the three trials, and Table 6 presents the same results for the bias indices.

Insert Tables 5 & 6 about here

All discrimination indices demonstrate a highly significant effect of trials by analysis of variance (all F's (2,200) > 237). Planned comparisons show that discrimination increases for all indices between trials 1 and 2 andfails to decrease between trial 2 and the delay trial.

As well, all bias indices change significantly across trials (all F's(2,200) > 9.34, all p's < .0001). Inspection of Table 6 suggests that normal subjects demonstrate conservative response biases on trial 1 which become close to neutral on trial 2 and slightly liberal by trial 3. Planned comparisons show that this change from conservative to neutral is significant for both bias indices between trials 1 and 2, but not between trial 2 and the delay trial for either bias index.

Patients' performance. Tables 7 and 8 present the preferred discrimination and bias indices for patients as a function of trial. In the one-way repeated measures analyses of variance performed, a conservative approach was taken, in which replications within subjects were collapsed to yield a single estimate for each of the 11 demented patients. This procedure yields a sample size representative of those used in studies with impaired populations.

Insert Tables 7 and 8 about here

Both d' and P_r show significant effects of trial, $F(3,30) = 7.23$, and $F(3,30) = 5.87$ respectively, both p's <.01. Planned comparisons were performed between trials 1 and 2, 2 and 3 and 3 and the delay, with identical results for each dependent measure. For both, there were significant increases between trials 1 and 2 [for d': $F(1,30) = 5.95$, $p < .05$; for P_r: $F(1,30) = 5.84$, $p < .05$], but not between trial 2 and trial 3, [for d': $F(1,30) = 3.72$, for P_r: $F(1,30) = 1.82$]. As well, neither measure showed forgetting across the delay [both F's < 1].

The measure B_r shows a significant increase in liberality across trials, [$F(3,30) = 4.37$, $p = .012$]. No pairwise sequential planned comparisons were significant [all p's > .05]. The decrease in C across trials fails to reach significance [$F(3,30) = 2.34$, $p = .09$]. Thus, in contrast to the discrimination measures, only B_r is sensitive to trials in showing that the apparent change toward more liberal bias in the patients is reliable.

Comparison of Patient and Normal Performance

In order to compare patients with normal subjects, we deleted trial 3 from the patient data set and performed 2 x 3 mixed design analyses of variance on the preferred discrimination and bias indices, with the between subjects variable diagnosis (demented versus normal) and the within subjects variable trial. In order to increase the power of the comparison, we selected the first administration of each form to each patient, and treated it as a separate subject in the analyses. This procedure generates 47 demented "subjects", compared to 101 normal subjects. Figures 7 and 8 show values of discrimination and bias indices for both groups of subjects as a *function of trials*.

Insert Figs 7 & 8 about here

As is apparent in Figure 7, normals perform better than patients on both discrimination measures, there is a significant effect of trials, and a significant interaction between trials and diagnosis due to the faster learning of the normals (all $\underline{p}$'s < .001).

More interesting are the bias differences between the two subject groups shown in Figure 8. For both B_r and C, there is a main effect of diagnostic group, with the demented subjects showing much more liberal biases than the normals [for B_r: $\underline{F}(1,146) = 27.86$; for C: $\underline{F}(1,146) = 45.44$, both $\underline{p}$'s < .0001]. Additionally the main effect for trials is significant for both bias measures, [for B_r: $\underline{F}(2,292) = 22.81$; for C: $\underline{F}(2,292) = 23.35$, both $\underline{p}$'s < .0001]. The interactions between diagnosis and trial are not significant for either measure [both $\underline{F}$'s < 1]. Furthermore, while the normals show slightly conservative biases on average, the demented patients show markedly liberal biases.

At the outset, we must admit that we did not succeed in equating *discrimination* between demented and normal subjects by increasing the numbers of items in the old and new sets. Not only were normals superior to demented subjects on the first trial, but they showed more rapid learning. Of more interest is the difference in the behaviors of the bias indices, with the demented subjects showing a robustly liberal bias, and the normals demonstrating an overall slightly conservative bias. Both groups become more liberal over trials in a parallel fashion.

Experiment 3

As Experiment 3, we report data from a study conducted by Butters, et al. (1985) who used a similar yes/no recognition memory task as that reported in Experiment 2 using words rather than pictures. They compared the memory performance of patients with Huntington's disease (HD) and amnesia (Am) with that of elderly normal controls. (We wish to thank them for providing us with their raw data for this analysis.)

We examined these data to extend the work reported in Experiment 2 to see if the differences we see between normal and abnormal recognition memory in Alzheimer's disease and parkinsonian dementia obtain in Huntington's disease (another dementing illness) with different stimulus material, and to extend our analysis of discrimination and bias to amnesic patients.

Method

Complete descriptions of the subjects, materials and procedure are given in Butters, et al (1985) and are briefly reviewed here.

Subjects and design. Subjects were 9 amnesics of various etiologies (6 Korsakoff's amnesics, 2 post encephalitic patients and 1 patient with a medialtemporal neoplasm), 10 patients with Huntington's disease with a range of functional disability, and 14 age-matched normal controls. The amnesics were older than the HD patients. The experiment followed a 3 (diagnosis: amnesic, demented, and normal) by 6 (test trial) mixed design.

Materials. Two forms of a 15 target - 15 distractor yes/no recognition task for words were constructed based on the RAVLT (Lezak, 1983). Additional items to serve as distractors were chosen to match the target items in frequency.

Procedure. The procedure used in this study was identical to that of Experiment 2 with the following exceptions. Five immediate study-test trials were administered; a 10 sec rest period was interposed between a test trial and the subsequent study trial; a 20 min filled delay intervened between the fifth test trial and the sixth (delayed) test trial.

Results and Discussion

Preferred discrimination and bias indices were computed as in Experiment 2. Discrimination indices averaged across trials for the three diagnostic groups are presented in Table 9; Table 10 gives the corresponding bias indices. These data were analyzed by mixed design 3 (diagnostic groups) x 6 (trials) analyses of variance.

Insert Tables 9 & 10 about here

Discrimination is best for normal subjects, worst for amnesics and intermediate for the Huntington's patients. Both discrimination indices show a significant main effect of subject group [for Pr: $F(2,30) = 50.74$; for d' : $F(2,20) = 71.56$, both p's < .0001] All the planned pairwise comparisons are significant, showing that normals are significantly better than Huntington's patients, who in turn are significantly better than amnesics.

Both measures also show a highly significant effect of trials [for Pr: $F(5,150) = 52.43$; for d': $F(5,150) = 50.43$, both p's < .0001]. The interaction between diagnosis and trials is significant for Pr, [$F(10,150) = 3.64$, $p < .001$], but not for d', [$F(10,150) = 1.57$, n.s.]. This interaction, shown in Figure 9, appears due to differential rates of forgetting between trial 5 and the delay trial among the groups -- the amnesics forget the most, the Huntington's patients forget a little, and the normals forget not at all. Although d' also shows the same pattern, the interaction for d' failed to reach significance.

Ins. Fig. 9 about here

As shown in Figure 10, the bias measures show a completely different pattern of results. The normal and amnesic subjects are somewhat conservative and do not appear to differ from one another, while the Huntington's patients show much more liberal bias than the other two groups.

Insert Fig. 10 about here

For Br, the effect of diagnostic group was significant [$F(2,30) = 3.73$, $p = .036$], with normals and amnesics showing somewhat conservative bias, and Huntington's patients showing liberal bias. There was a highly significant effect of trials, with bias becoming more liberal across trials [$F(5,150) = 9.40$, $p < .0001$]. The interaction was not significant [$F<1$]. For C, the differences among diagnostic groups failed to reach significance [$F(2,30) = 3.00$, $p = .065$]. There was, however, a highly significant effect of trials [$F(5,150) = 9.45$, $p < .0001$], and no interaction [$F<1$].

To recapitulate, both models show the following important results: in terms of discrimination, amnesic memory is worse than demented memory which in turn is worse than normal memory. The effect of trials was *significant for both discrimination and bias*. All groups learned (that is, showed increases in their discrimination indices), and all groups became more liberal over trials. More interestingly, demented bias is more liberal than either amnesic or normal bias, which are equal. This last finding replicates our own results from Experiment 2 and a number of similar observations in the literature. We take these particular bias results quite seriously, since they seem to indicate that liberal bias is not solely a result of poor memory (or the amnesic patients would surely show it too). Thus, there is some property of dementia which causes new items to appear familiar, or old.

Turning next to the more subtle differences in patterns of results for the two preferred models, we found that although both discrimination indices show a significant effect of group and of trials, only Pr shows a significant interaction, indicating that normal and demented subjects learn at a greater rate and forget less than the amnesics.

In addition, although both bias measures show the expected difference between demented subjects, who show liberal response bias, and amnesics and normals, who show neutral or somewhat conservative response bias, only B_r shows a significant increase in bias with trials (a phenomenon also observed in Experiment 2 and again only significantly for B_r).

General Discussion

There are several major findings from these explorations of recognition memory. First, not all commonly used theories of the recognition memory process fare equally well in yielding independence of discrimination from response bias. The two models which pass the independence test are two-high threshold and signal detection theory with the intersection bias index C rather than ß, the ratio bias index.

Secondly, the models appear differentially sensitive to changes in bias and discrimination. Across the patient groups, the discrimination index P_r derived from two-high threshold theory was not as sensitive to change in performance as the d' measure from SDT; but the SDT bias measure C was not as sensitive to change within the AD-PD group as the two-high threshold bias measure B_r.

Thirdly, all subjects, normal, demented and amnesic, demonstrated parallel increases in liberality of response bias across learning trials.

Finally, normal response bias distinguishes amnesia from the three types of dementia tested. Alzheimer's, parkinsonian and Huntington's dementia were all marked by abnormally liberal bias in addition to poor discrimination.

The following discussion addresses the implications of each of the major findings in turn.

Acceptability of Models

Only two models of recognition memory yield indices of discrimination which are independent of bias. The acceptable theories are two-high threshold with P_r and B_r, and SDT with normal and logistic distributions using d' and C.

When other indices are reported, experimenters are faced with such tasks as *disambiguating changes in the* bias index which appear as functions of changes in discrimination (e.g.: as with d' and ß), or explaining changes in a bias index which is invalid under prevailing experimental conditions (e.g. as with C_j in repeated study-test paradigms) (for examples see Hart, Smith & Swash, 1985; Mohs & Davis, 1982). Discussion of the meaning of results is simplified by use of either of the two models in which the experimenter is assured that memory and response strategy measures are independent.

Differential Sensitivity of Models

At first glance, it appears that d' is more sensitive to learning and forgetting than P_r, and therefore we should recommend use of SDT models to examine small changes in discrimination in both normal and impaired subjects. However, the opposite pattern of sensitivity seems to hold for the bias measures in Experiment 2 with AD and PD patients. That is, significantly increased liberality of bias between trial 1 and trial 2 is demonstrated by B_r but not by C.

What should the researcher do? As before, we believe the answer is a pragmatic one. We recommend the use of both acceptable models to characterize recognition memory under any manipulation. A negative result under one theory may be positive under another; both should be reported and discussed. If multiple non-independent comparisons are problematic, stricter criteria for excluding chance significance may be used. Further, if the experimental manipulation is designed to improve performance to a clinically relevant degree (as in trials of novel therapeutic agents for AD), efficacy is best supported if improvement in seen in both two-high threshold and SDT measures.

Increasing Liberality Over Trials

All subjects in Experiment 2, regardless of diagnosis, demonstrated increased liberality of response bias across learning trials in the absence of explicit manipulations of payoff matrix. This effect was most prominent between the first and second learning trials. Mohs and Davis' (1982) AD subjects also showed this effect in C and B_r as we calculated them from the presented raw data. The simplest explanation of this universal increase in liberality across learning trials is a buildup of interference across repeated study-test trials. That is, as more items are presented, the net familiarity of all items shifts upwards (in SDT terms), while these subjects' absolute level of familiarity used as the criterion value remains fixed. This leads to an overall increase in both hits and false alarms without an actual downward shift of criterion. The liberality effect is more difficult to explain in two-high threshold terms. In this case, the theory holds that increases in hits and false alarms are a result of increased guesses of yes when uncertain. That is, there must be a real change in response bias, rather than a simple upward shift in net familiarity. Thus, the SDT model is a more appealing conceptualization.

Response Bias and Diagnosis

Why might amnesic subjects have normal reponse bias in the face of the worst discrimination performance of all patients tested? Alternatively phrased, why do the demented subjects show the most liberal response bias when their discrimination is intermediate between normal and amnesic subjects?

Two major explanatory systems can be brought to bear on this question. A neurochemical explanation would focus on the neurotransmitter deficits characteristic of each disorder, and conclude that these differences sufficiently accounted for the different patterns of memory deficit. Cognitive psychological explanations would focus on the psychological processes which differed between groups. Ideally, the two types of explanation would coincide such that the learning process differences would be mediated by the neurotransmitter systems damaged in each disorder.

We turn first to a cognitive psychological explanation of the recognition memory process. A feature match model of recognition holds that recognition takes place via pattern matching between features of the test stimuli and stored active representations of the "old" items. If a match or near match is found, the subject responds "yes", otherwise the subject responds "no". In terms of familiarity, an increasing number of features in common between the test items and internal representations (either from the target set or from semantic memory) yields a parallel increase in subjective familiarity. Thus, subjects might have high levels of familiarity for both old and new items if stimuli are not encoded distinctively at presentation and semantic memory representations have been inappropriately activated by prior target and distractor items.

Thus, stimuli which are not encoded distinctively at presentation will adequately match stored representations of similiarly poorly encoded target items or very familiar items in semantic memory, leading to a general increase in "yes" responses to both old and new items.

There is some evidence for this "distinctiveness" hypothesis of failure to distinguish between old and new items. It has been found repeatedly that AD patients perform poorly on tasks such as category retrieval, object naming and similarities (e.g., Bayles, 1982; Ober, Dronkers, Koss, Delis, & Friedland, 1986). The most common interpretations of these results is that AD produces disorders in semantic memory structure or access (Martin & Fedio, 1983; Nebes, Boller, & Holland, 1986; Ober et al, 1986; Warrington, 1975).

Two sets of findings are particularly germane. The most direct demonstration of failure to encode distinctive features has been provided by Grober, Buschke, Kawas, and Fuld (in press). These investigators asked mildly impaired AD subjects and matched normal controls to rate features of concepts, and found that AD patients did not distinguish as well between prototypical features (e.g.: LEGS - CHAIR) and non-critical features (e.g.: CLOTH - CHAIR). Additionally, Nebes and colleagues (1986) have shown that AD patients are more sensitive to context effects on a sentence completion task in that these patients performed quite poorly unless the sentence frame highly constrained the set of potential responses. These results may be interpreted as a consequence of failure to locate appropriate features of potential responses given less than optimal search cues.

On the other hand, normal semantic memory and normal intelligence are defining features of amnesia (Moskovitch, 1982). Therefore, it is not surprising that priming, category retrieval and similarities tasks are performed near normally by amnesics (Parkin, 1984).

Thus, whether at encoding, storage or retrieval, AD patients show evidence of blurring of distinctive features of memoranda; to our knowledge, this dysfunction has not been reported in amnesia.

Neurochemically, a case has been made for the importance of cortical acetylcholine deficits in the production of abnormally liberal bias in recognition and intrusion errors in recall tasks (Fuld, Katzman, Davies, & Terry, 1982; Mohs & Davis, 1982). Mohs and Davis' AD patients did demonstrate a reduction in liberality of bias with physostigmine treatment (which enhances synaptic acetylcholine availability) evident in both C and B_r as calculated by the present authors. Their experimental design does not allow one to say that this treatment yielded a near-normal pattern of bias because there were no age-matched normal control subjects, but the changes seen were in the desired direction.

However, abnormally liberal bias is evident as well in the Butters' et al. sample of Huntington's disease patients in whom cortical cholinergic deficits are not prominent (Pearce, Sofroniew, Cuello, Powell, Eckenstein, Esiri, & Davison, 1984). Thus, although the dementia deficit in bias is clear, its neurochemistry remains to be elucidated.

SUMMARY

In summary, we would like to reiterate the following points:

1. Any pair of discrimination and bias indices assumes a theory of the recognition memory process. There are no theory-free measures of performance.

2. Abnormal bias is functionally as important as abnormal discrimination.

3. Only two-high threshold and signal detection theory with C yield measures of discrimination and bias which are independent.

4. These two models appear differentially sensitive to changes in bias and discrimination. Therefore, we suggest using both to characterize recognition memory performance, especially in trials of novel therapies.

5. Patients with Huntington's, Alzheimer's and parkinsonian dementia all demonstrated abnormally liberal bias in addition to poor discrimination. The sample of amnesic patients tested had normal bias, despite very poor discrimination.

6. All subjects, regardless of diagnosis, become increasingly liberal over repeated trials without explicit payoff manipulations.

References

American Psychiatric Association (1980). *Diagnostic and Statistical Manual of Mental Disorders*, third edition *(DSMIII)*. Washington, D.C.: Author.

Banks, W. P. (1970). Signal detection theory and human memory. *Psychological Bulletin, 74*, 81-99.

Bartus, R.T., Dean, R.L., Beer, B. , & Lippa, A.S. (1982) . The cholinergic hypothesis of geriatric memory dysfunction. *Science , 217*, 408-417.

Battig , W.F., & Montague , W.E. (1969). Category norms for verbal items in fifty-six categories: a replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monograph 80 (3, pt. 4)*.

Bayles, K.A. (1982). Language function in senile dementia. *Brain and Language, 16*, 265-280.

Branconnier, R. J., Cole, J. O., Spera, K. F., & DeVitt, D. R. (1982). Recall and recognition as diagnostic indices of malignant memory loss in senile dementia: A Bayesian analysis. *Experimental Aging Research, 8*, 189-193.

Brinkman , S.D., & Gershon , S. (1983). Measurement of cholinergic drug effects on memory in Alzheimer's disease. *Neurobiology of Aging ,4*, 139-145.

Butters, N., Wolfe, J., Martone, M., Granholm, E., & Cermak, L. S. (1985). Memory disorders associated with Huntington's disease: Verbal recall, verbal recognition and procedural memory. *Neuropsychologia, 23*, 729-743.

Crook, T., Gilbert, J.G., & Ferris, S. (1980). Operationalizing memory impairment for elderly persons: The Guild Memory Test. *Psychological Reports, 47*, 1316-1318.

Egan, J. P. (1958). *Recognition memory and the operating characteristic*. (Tech. Note). Bloomington, IN: Indiana University Hearing and Communication Laboratory.

Fuld, P.A., Katzman, R., Davies, P., & Terry, R. D. (1982). Intrusions as a sign of Alzheimer's disease: Chemical and pathological verification. *Annals of Neurology, 11*, 155-159.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91*, 1-67.

Glanzer, M. & Adams, J. (1985). The mirror effect in recognition memory. *Memory and Cognition 13*, 8-20.

Green, D.M., & Moses, F. L (1966). On the equivalence of two recognition measures of short-term memory. *Psychological Bulletin, 66*, 228-234.

Green, D.M., & Swets, J.A. (1974). *Signal detection theory and psychophysics.* Huntington, NY: Krieger.

Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin, 75*, 424-429.

Grober, E., Buschke, H., Kawas, C., & Fuld, P. (in press). Impaired ranking of semantic attributes in dementia. *Brain and Language.*

Hart, S. A., Smith, C. M., & Swash, M. (1985). Recognition memory in Alzheimer's disease. *Neurobiology of Aging, 6*, 287-292.

Hodos, W. (1970). A nonparametric index of response bias for use in detection and recognition experiments. *Psychological Bulletin, 74*, 351-354.

Lezak, M. (1983). *Neuropsychological assessment* (2nd Ed.). New York: Oxford.

Link, S.W., & Heath, R.A. (1975). A sequential theory of psychological discrimination. *Psychometrika, 40*, 77-105.

Lockhart, R.S., & Murdock, B.B., Jr. (1970). Memory and the theory of signal detection. *Psychological Bulletin, 74*, 100-109.

Luce, R.D. (1959). *Individual choice behavior.* New York: Wiley.

Luce, R.D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology (Vol. I).* New York: Wiley.

Martin, A. & Fedio, P. (1983). Word production and comprehension in Alzheimer's disease: The breakdown of semantic knowledge. *Brain and Language, 19*, 124-141.

McKhann, G., Drachman, D., Folstein M., & Katzman, R. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA work group. *Neurology, 34*, 939-944.

Mohs, R.C., & Davis, K. L. (1982). A signal detectability analysis of the effect of physostigmine on memory in patients with Alzheimer's disease. *Neurobiology of Aging, 3*, 105-110.

Moskovitch, M. (1982). Multiple dissociations of function in amnesia. In L.S. Cermak (Ed.) *Human memory and amnesia* (Pp. 337-370). Hillsdale, NJ:Erlbaum,.

Murdock, B.B., Jr. (1974). *Human memory: Theory and data.* Potomac, MD: Erlbaum.

Nebes, R.D., Boller, F. & Holland, A. (1986). Use of semantic context by patients with Alzheimer's disease. *Psychology and Aging, 3*, 261-269.

Noreen, D. L. (1977). Relations among some models of choice. Paper presented at the Tenth Annual Mathematical Psychology Meetings, University of California, Los Angeles, California, August.

Ober, B.A., Dronkers, N.F., Koss, E., Delis, D.C., & Friedland, R.P. (1986). Retrieval from semantic memory in Alzheimer-type dementia. *Journal of Clinical and Experimental Neuropsychology, 8,* 75-92.

Paivio, A., Yuille, J.C., & Madigan, S. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology, 76* (Monograph Suppl.), No. 1, Pt. 2.

Parkin A. (1984). Amnesic syndrome: A lesion-specific disorder? *Cortex, 20,* 479-508.

Pearce, R.C.A., Sofroniew, M.V., Cuello, A.C., Powell, T.P.S., Eckenstein, R., Esiri, M.M., & Davison, A.N. (1984). Neurotransmitter dysfunction and atrophy of the caudate nucleus in Alzheimer's disease. *Neurochemical Pathology, 2,* 221-233.

Pollack, I., & Norman, D.A. (1964). A nonparametric analysis of recognition experiments. *Psychonomic Science, 1,* 125-126.

Reisberg, B., Ferris, S.H., & Crook, T. (1982). Signs, symptoms, and course of age-related cognitive decline. In S. Corkin, K.L. Davis, J.H. Growdon, E. Usdin and R.J. Wurtman. (Eds.), *Alzheimer's disease: A report of progress in research (Aging,* Vol. 19, Pp. 177-181). New York: Raven Press.

Semple, S.A., Smith , C.A. & Swash , M. (1982). The Alzheimer disease syndrome. In S. Corkin, K.L. Davis, J.H. Growdon, E. Usdin and R.J. Wurtman (Eds.) *Alzheimer's disease: A report of progress in research (Aging,* Vol. 19, pp. 93-108). New York: Raven Press.

Snodgrass, J. G. (1983). The pragmatics of measuring recognition memory performance. Paper presented at the meetings of the Psychonomic Society, San Antonio, TX.

Snodgrass , J.G. & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, familiarity and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6* , 174-215.

Thorndike, E.L., & Lorge, I. (1944). The teacher's word book of 30,000 words. New York: Columbia University, Teacher's College.

Tulving, E., & Thomson, D. M. (1971). Retrieval processes in recognition memory: Effects of associative context. *Journal of Experimental Psychology,* 87, 116-124.

Upton, G. J. G. (1978). *The analysis of cross-tabulated data.* Chichester, England: Wiley.

Warrington, E. K. (1975). The selective impairment of semantic memory. *Quarterly Journal of Experimental Psychology, 27,* 635-657.

Woodworth, R.S. (1938). *Experimental psychology.* New York: Holt.

Footnote

The authors would like to thank Nelson Butters, Eric Granholm and their colleagues for making their raw data available. Kelly Feenan's help in running subjects and analyzing data is greatly appreciated. Preliminary versions of this paper were presented at the 1983 meeting of the Psychonomics Society in San Antonio, Texas and at the 1986 meeting of the Eastern Psychological Association in New York. This research was supported in part by a Research Challenge Fund grant from New York University, a University Research Initiative Program grant from the Air Force Office of Scientific Research, and a Merit Review Program grant from the Veterans Administration.

FIGURE CAPTIONS

Figure 1. Stimulus-response matrix (A) and corrected matrix (B) illustrating the correction for computing hit and false alarm rates.

Figure 2. Definitions of C (part A) and ß (part B) in the signal detection model.

Figure 3. Areas in the unit square used to define the nonparametric indices.

Figure 4. Iso-memory and Iso-bias functions for four models of recognition memory.

Figure 5. Relationship between bias and discrimination for four models of recognition memory. (N.B.: Open circles represent the maximum possible conservative bias; closed circles represent the maximum possible liberal bias.)

Figure 6. Hit and false alarm rates for Experiment 1.

Figure 7. Discrimination values (P_r and d') for normal and demented subjects by trials for Experiment 2.

Figure 8. Bias values (B_r and C) for normal and demented subjects by trials for Experiment 2.

Figure 9. Discrimination indices (P_r and d') by trial for Butters' subjects.

Figure 10. Bias indices (B_r and C) by trial for Butters' subjects.

Table 1

Summary of discrimination and bias measures

A. SDT with normal distributions

(1) $d' = z_{FA} - z_H$

(2) $\ln(\beta) = \ln[f_o(z_H)/f_n(z_{FA})]$

(3) $C = z_{FA} - d'/2 = 0.5\,(z_{FA} + z_H)$

B. SDT with logistic distributions

(4) $d_L = \ln\{[H(1\text{-}FA)]/[(1\text{-}H)FA]\}$

(5) $\ln(\beta_L) = \ln\{[H(1\text{-}H)]/[FA(1\text{-}FA)]\}$

(6) $C_L = 0.5\,[\ln\{[(1\text{-}FA)(1\text{-}H)]/[(H)(FA)]\}]$

C. Two-high-threshold theory

(7) $P_r = H - FA$

(8) $B_r = FA/[1-(H-FA)]$

D. Distribution free (nonparametric) theory

For H ≥ FA:	(9a)	$A' = 0.5 + [(H\text{-}FA)(1+H\text{-}FA)]/[4H(1\text{-}FA)]$
For FA > H:	(9b)	$A' = 0.5 - [(FA\text{-}H)(1+FA\text{-}H)]/[4FA(1\text{-}H)]$
For H ≥ FA:	(10a)	$B'' = [H(1\text{-}H)\text{-}FA(1\text{-}FA)]/[H(1\text{-}H)+FA(1\text{-}FA)]$
For FA > H:	(10b)	$B'' = [FA(1\text{-}FA)\text{-}H(1\text{-}H)]/[FA(1\text{-}FA)+H(1\text{-}H)]$
For H ≤ (1-FA):	(11a)	$B'_H = 1 - \{[FA(1\text{-}FA)]/[H(1\text{-}H)]\}$
For H > (1-FA):	(11b)	$B'_H = \{[H(1\text{-}H)]/[FA(1\text{-}FA)]\} - 1$

Note: H = (#hits+0.5)/(#olds +1); FA = (#fas +0.5)/(#olds +1)

Table 2

Iso-memory and iso-bias functions for four models of recognition memory.

A. SDT with logistic distributions and β_L

Iso-memory: $H = aFA/[(a-1)FA + 1]$, where $a = e^{d_L}$

Iso-bias for β_L: $H = \{1+[1+4\beta_L(FA^2-FA)]^{0.5}\}/2$

B. SDT with logistic distributions and C_L

Iso-memory: $H = aFA/[(a-1)FA + 1]$, where $a = e^{d_L}$

Iso-bias for C_L: $H = (1-FA)/(aFA-FA+1)$ where $a=e^{2C_L}$

C. Two-high threshold theory

Iso-memory: $H = P_r + FA$

Iso-bas: $H = [(B_r-1)/B_r]FA + 1$

D. Distribution free (non-parametric) theory with B"

Iso-memory: $H = \min(1, [FA(1-FA)+k^2]^{0.5} - k$,

where $k = 1.5-2[FA+A'(1-FA)]$

Iso-bias with B": $H = 0.5 \pm \{0.25-[FA(1-FA)(1+B")]/(1-B")\}^{0.5}$

Table 3

Mean values of discrimination indices for the four basic models

in Experiment 1.

Imagery	*high imagery*			*low imagery*		
Payoff	*cons*	*neut*	*lib*	*cons*	*neut*	*lib*
*~Hits	.716	.826	.868	.474	.632	.723
*~False Alarms	.097	.161	.219	.203	.303	.397
~~d'(SDT:norm)	2.013	2.090	2.148	.818	.897	.935
~~d'(SDT:log)	3.480	3.580	3.738	1.381	1.467	1.553
~~P_r(2HT)	.619	.664	.648	.271	.329	.326
~~A'(nonpar)	.880	.895	.886	.708	.731	.724

*~significant effect of payoff, no interaction

~~neither payoff nor interaction significant

n.b. all measures showed a significant effect of imagery

Table 4

Mean values of bias indices for the four basic models in Experiment 1.

Payoff	*conservative*		*neutral*		*liberal*	
Imagery	*high*	*low*	*high*	*low*	*high*	*low*
*~Hit rate	.716	.474	.826	.632	.868	.723
*~False alarm rate	.097	.203	.161	.303	.219	.397
~*ln ß (SDT:norm)	.635	.350	.077	.016	-.265	-.061
~~C (SDT:norm)	.363	.485	.032	.085	-.161	-.162
~*ln $ß_L$(SDT:log)	.777	.435	.093	.020	-.324	-.077
~~C_L(SDT:log)	.659	.816	.063	.137	-.290	-.265
~~B_r (2HT)	.275	.281	.475	.464	.597	.572
~*B" (nonpar)	.354	.224	.043	.019	-.148	-.041
~*B_H (nonpar)	.490	.285	.060	.015	-.210	-.073

*~imagery significant, no interaction

~*imagery n.s., interaction significant

~~neither significant

n.b. payoff is significant for all measures

Table 5

Mean values of discrimination indices for normal subjects

by test trial in Experiment 2

	Trial 1	Trial 2	Delay
Hit rate	.787	.952	.960
False alarm rate	.125	.041	.042
d' (SDT:norm)	2.09	3.70	3.76
P_r (2HT)	.662	.910	.917

N.B.: both discrimination measures show significant changes towards better discrimination between trials 1 and 2, but not between trial 2 and the delay trial.

Table 6

Mean values of bias indices for normal subjects

by test trial in Experiment 2

	Trial 1	Trial 2	Delay
Hit rate	.787	.952	.960
False alarm rate	.125	.041	.042
C (SDT:norm)	.180	.024	-.019
B_r (2HT)	.382	.487	.518

N.B.:. both bias indices show significant changes from conservative to neutral bias from trial 1 to trial 2, but no change between trial 2 and the delay trial.

Table 7

Mean values of discrimination indices for patients

by test trial in Experiment 2

	Trial 1	Trial 2	Trial 3	Delay
Hit rate	.782	.843	.897	.888
False alarm rate	.359	.326	.327	.330
d'(SDT:norm)	1.46	1.80	2.06	1.95
Pr(2HT)	.423	.517	.570	.558

N.B.: both discrimination measures show significant changes towards better discrimination between trials 1 and 2, but not between trial 2 and 3 or trial 3 and the delay.

Table 8

Mean values of bias indices for patients

by test trial in Experiment 2

	Trial 1	Trial 2	Trial 3	Delay
Hit rate	.782	.843	.897	.888
False alarm rate	.359	.326	.327	.330
C (SDT:norm)	-.221	-.256	-.372	-.388
B_r (2HT)	.537	.649	.773	.779

N.B.:. both bias indices show significant changes toward more liberal bias from trial 1 to trial 2, but no change between trial 2 and 3 or between trial 3 and the delay.

Table 9

Mean values of discrimination indices for the three groups of subjects in Butters et al. (1985)

	Normals	Amnesics	Huntington's
Hit rate	.859	.662	.804
False alarm rate	.043	.227	.210
d' (SDT:norm)	2.90	1.34	1.90
P_r(2HT)	.816	.434	.594

N.B.: Both discrimination measures show significant differences among groups.

Table 10

Mean values of bias indices for the three groups of subjects in Butters et al. (1985)

	Normals	Amnesics	Huntington's
Hit rate	.859	.662	.804
False alarm rate	.043	.227	.210
C (SDT:norm)	.320	.196	.014
B_r (2HT)	.243	.395	.474

N.B.: Of the two bias indices, only B_r shows a significant effect of groups, with C only marginally significant.

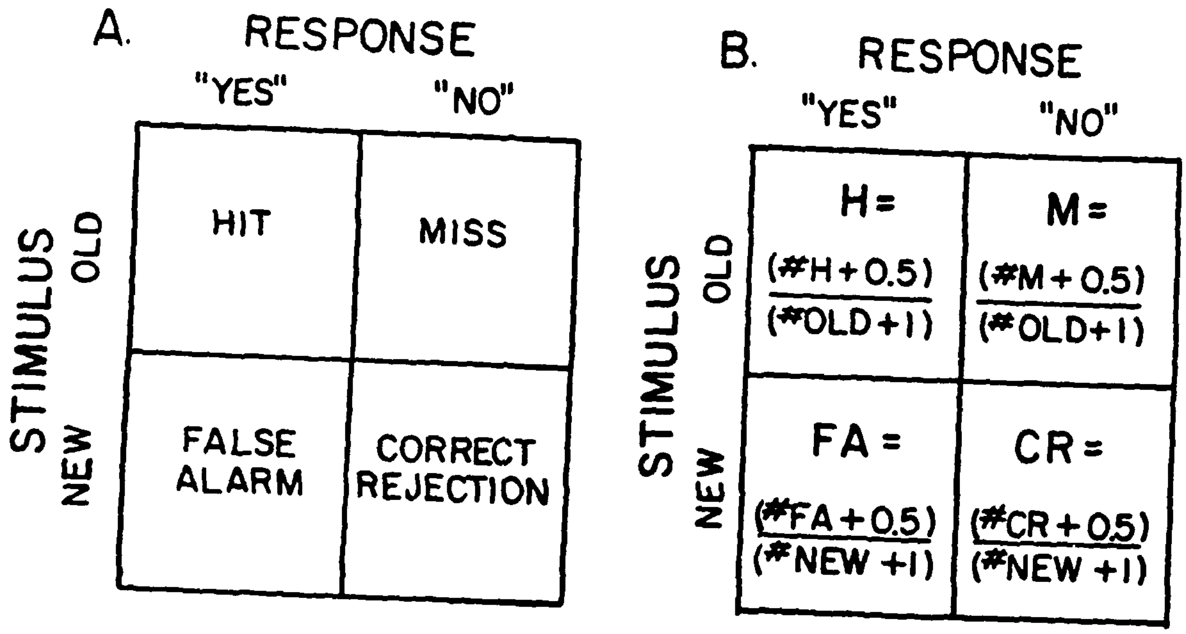


Figure 1. Stimulus-response matrix (A) and corrected matrix (B) illustrating the correction for computing hit and false alarm rates.

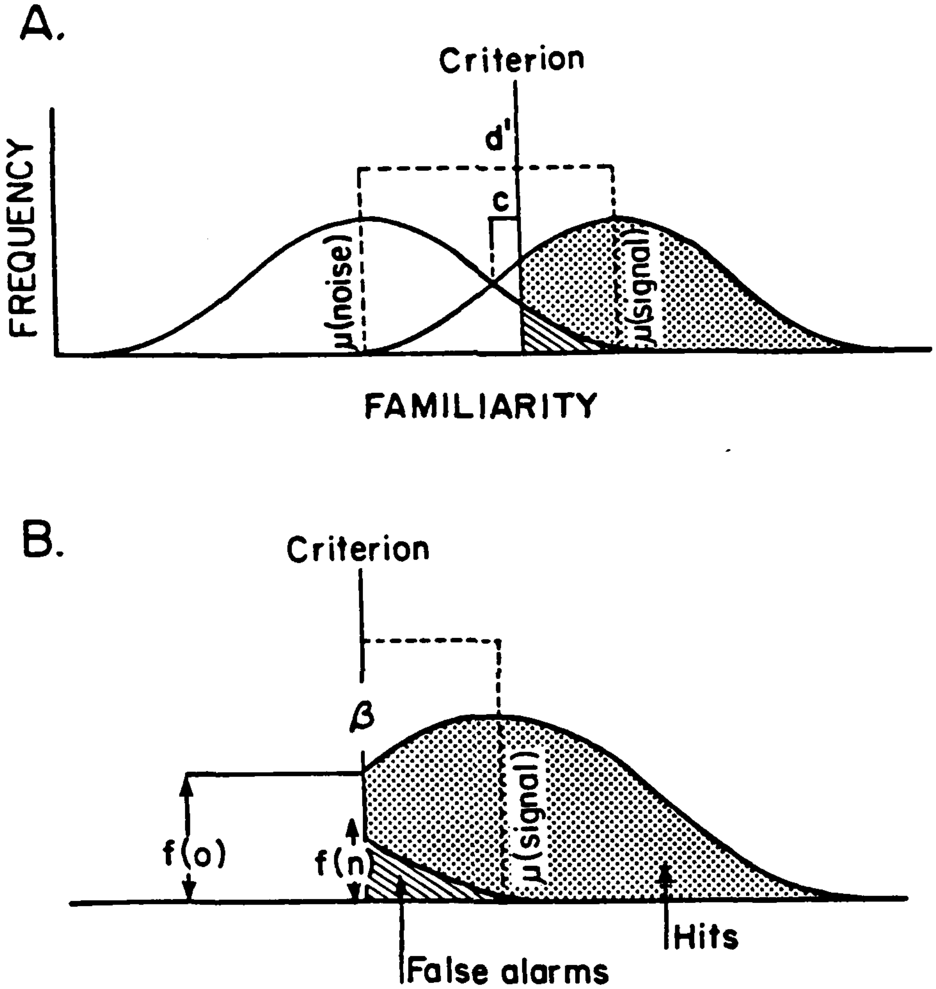


Figure 2. Definitions of C (part A) and β (part B) in the signal detection model.

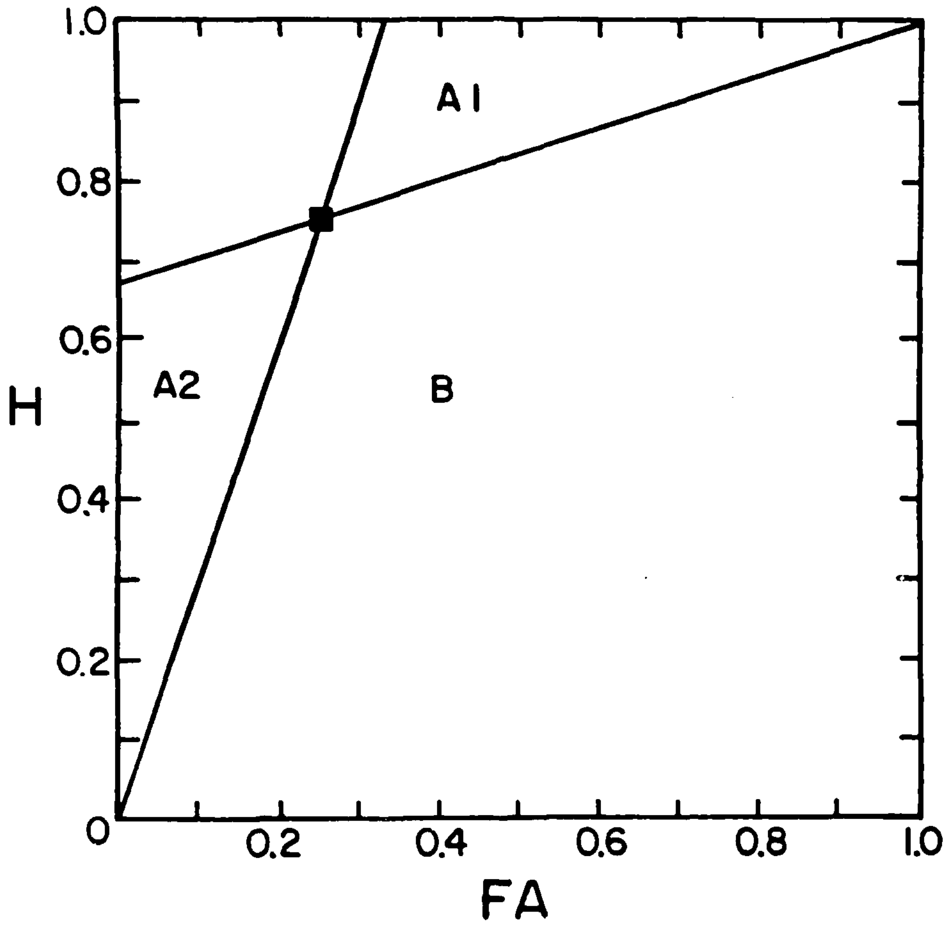


Figure 3. Areas in the unit square used to define the nonparametric indices.

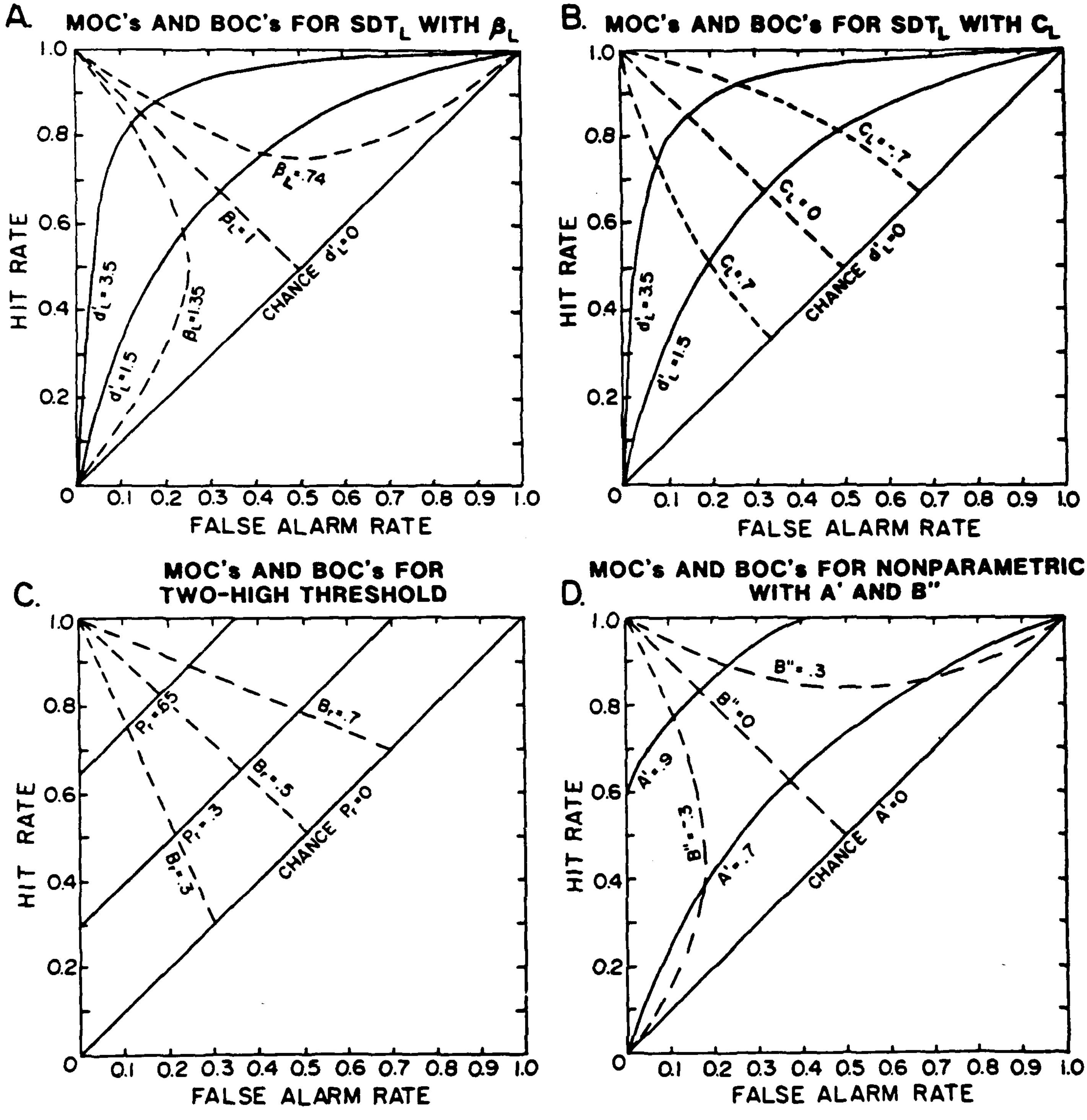


Figure 4. Iso-memory and iso-bias functions for four models of recognition memory.

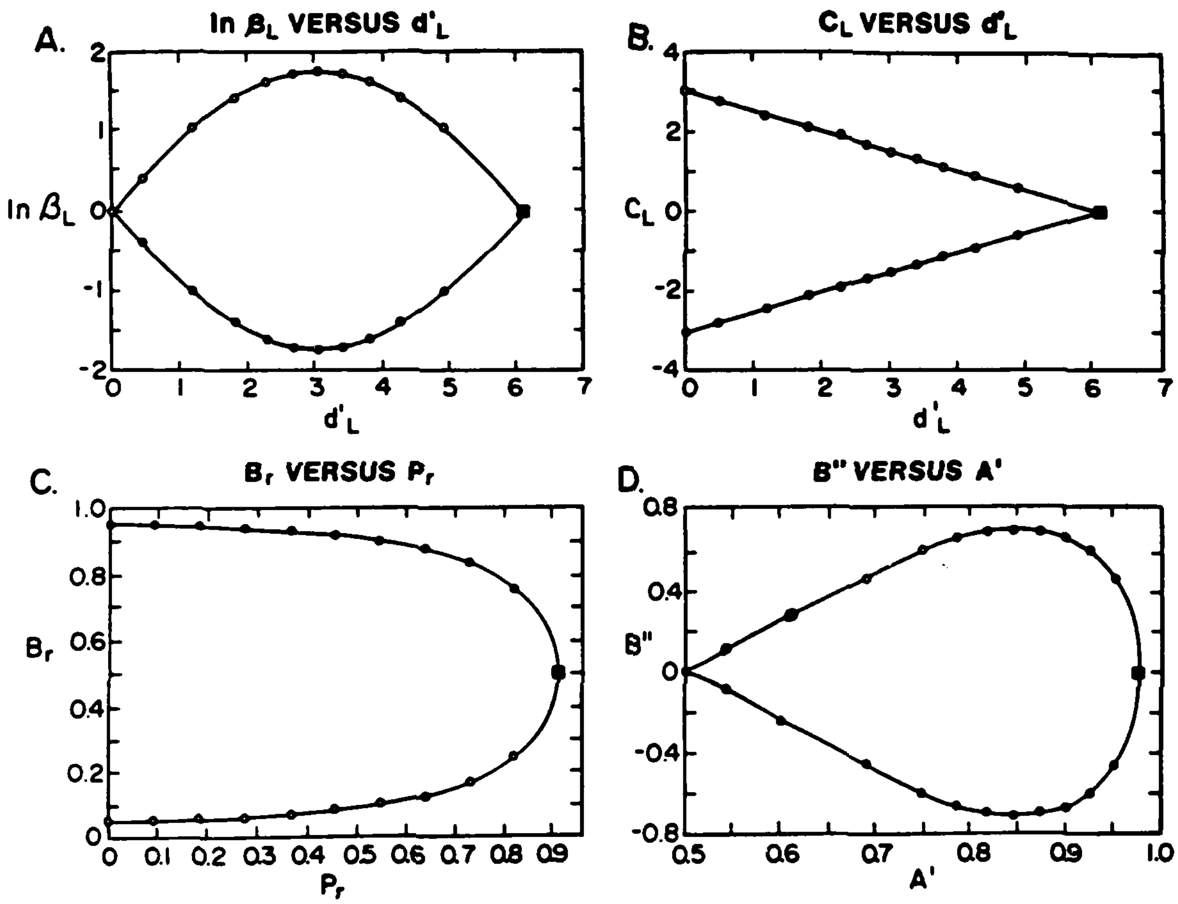


Figure 5. Relationship between bias and discrimination for four models of recognition memory. (N.B.: Open circles represent the maximum possible conservative bias; closed circles represent the *maximum* possible liberal bias.)

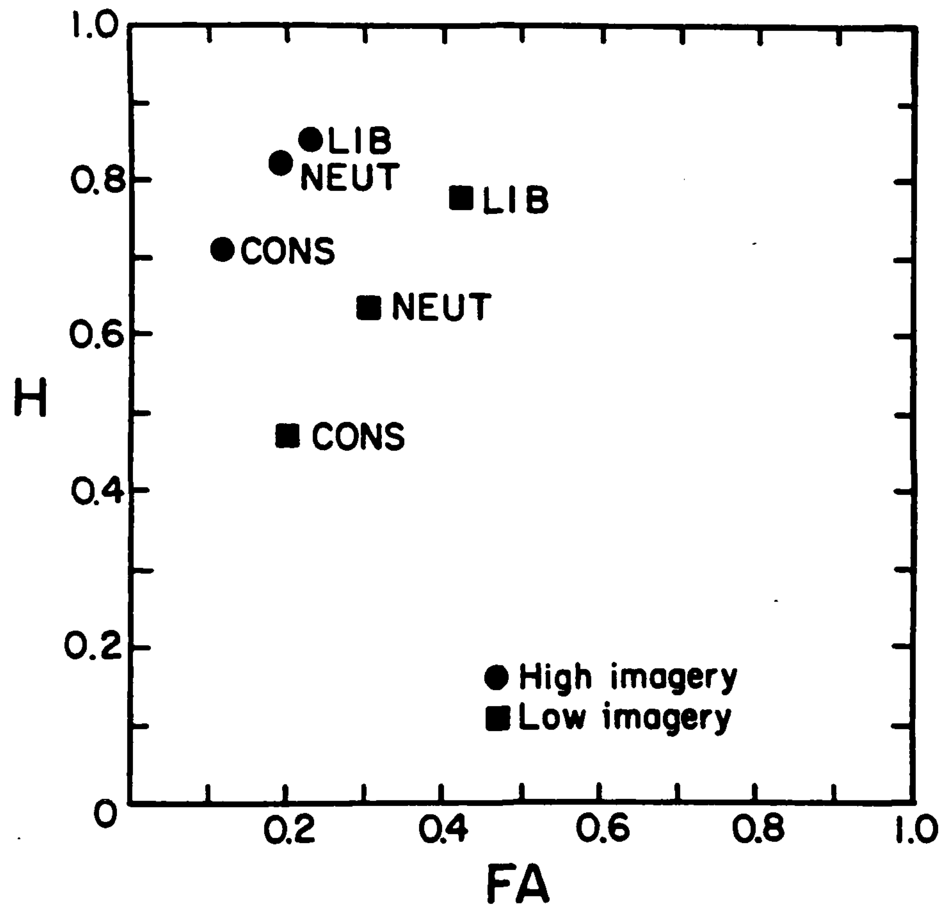


Figure 6. Hit and false alarm rates for Experiment 1.

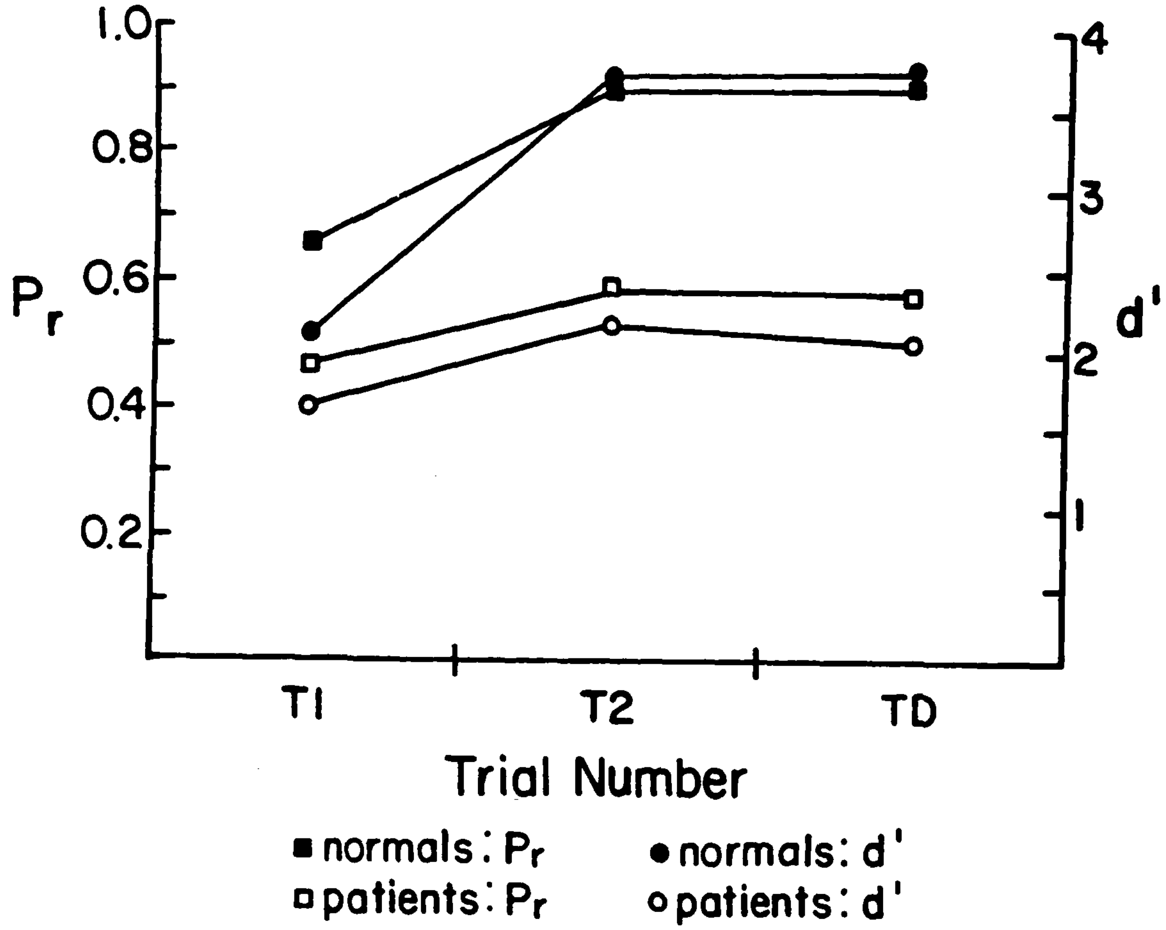


Figure 7. Discrimination values (P_r and C) for normal and demented subjects by trial for Experiment 2.

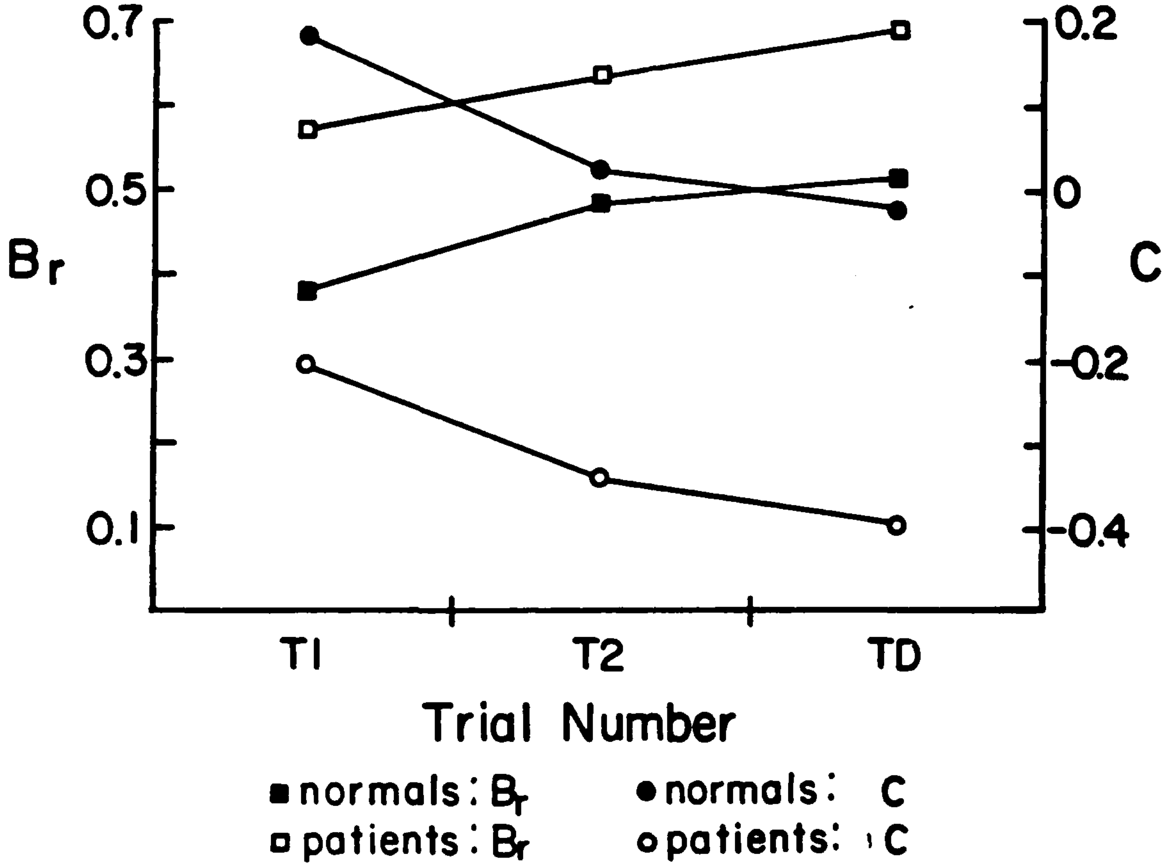


Figure 8. Bias values (B_r and C) for normal and demented subjects by trials for Experiment 2.

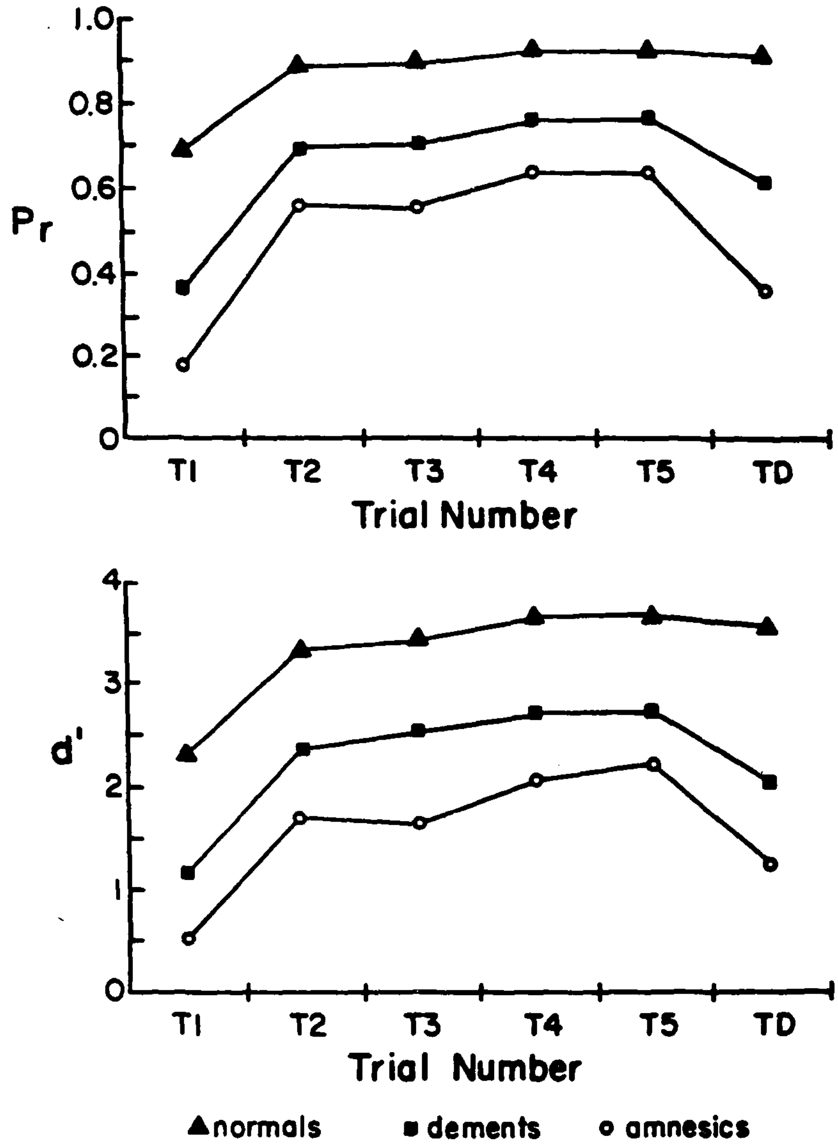


Figure 9. Discrimination indices (P_r and d') by trial for Butters' subjects.

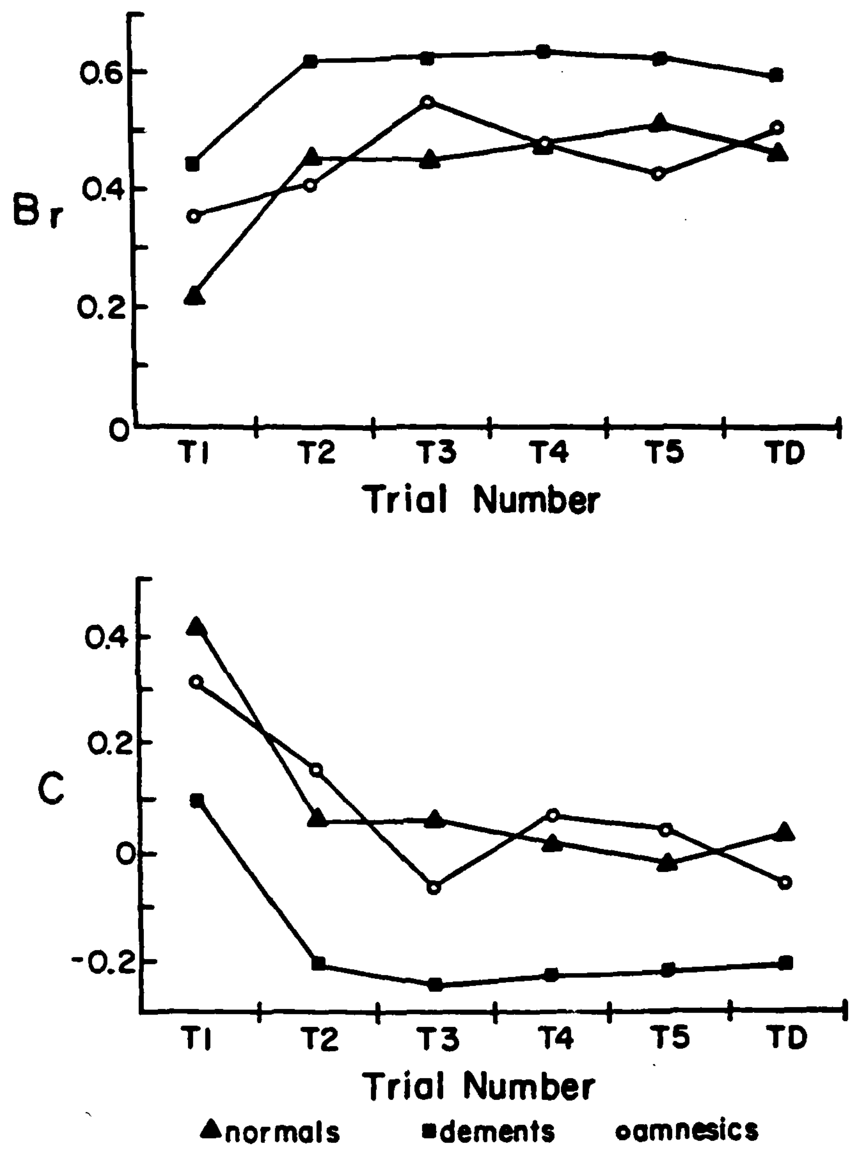


Figure 10. Bias indices (B_r and C) by trial for Butters' subjects.

Presented at the International Cryogenic Materials Conference
and Cryogenic Engineering Conference
at Fermi National Accelerator Laboratory, Batavia, Illinois
June 14-18, 1987

and to be published in

Advances in Cryogenic Engineering
Plenum Press, New York

INSTRUMENTATION FOR CLINICAL APPLICATIONS OF NEUROMAGNETISM

D.S. Buchanan, D. Paulson

Biomagnetic Technologies, inc.
San Diego, California

and

S.J. Williamson

Neuromagnetism Laboratory
Department of Physics
New York University, New York, New York

ABSTRACT

Measurements of the magnetic field of the human brain in a clinical setting require a higher level of performance from the instrumentation than is generally acceptable for laboratory research. We describe several significant advances that are intended for such neuromagnetic applications as well as for a broader range of biomagnetic studies. We report the development of a closed-cycle refrigerator capable of sustaining a SQUID-based sensor without introducing significant deterioration of the noise level. This eliminates the need for liquid cryogens and permits the sensor to be operated in various orientations, including inverted. The performance of a new type of magnetically shielded room is then evaluated for neuromagnetic studies. It has the advantages of being pre-fabricated and of providing a large interior for convenient clinical studies. Its ceiling supports a versatile gantry that holds one or two sets of magnetic sensors. This arrangement, when used with a magnetic system for precisely determining the sensor positions with respect to the patient's head, is feasible for precise localization of neural sources within the brain. We end with an example of the kinds of clinical studies that are now being carried out with the aid of neuromagnetic measurements.

INTRODUCTION

The technology for magnetic studies of the brain has greatly advanced since the first observation of neuromagnetic fields by Cohen.[1,2] It is now recognized that the dominant source of magnetic fields measured outside the scalp is the pattern of intracellular currents in active neurons. Since magnetic fields in the frequency domain of interest are little affected by the electrical properties of the cranium, they emerge from the head without distortion. One important advantage of magnetic studies is the possibility of determining the three-dimensional location of active neural regions in the brain through measurements and analysis of the field pattern measured over the scalp. These methods are described in recent reviews[3-5] and in a textbook.[6] In many cases it is possible to localize a confined source with a precision of a few millimeters. As the magnetic fields of interest range in strength from about 10 fT to 1000 fT, the only field sensor having the required sensitivity and small sampling volume is the SQUID (superconducting quantum interference device). The need for a dewar to contain the liquid helium that maintains the SQUID below its superconducting transition temperature imposes significant contraints on how and where such a sensor may be used.

The past three years have seen several multi-sensor systems developed for neuromagnetic applications.[7-9] One motivation for introducing an array of sensors within a single dewar is to speed neuromagnetic measurements by reducing the number of times the dewar must be moved from one location to another over the scalp to record a field pattern. While virtually all SQUIDs presently sense the fields of interest by means of a flux transformer connected to a detection coil wound of superconducting wire, considerable effort is being invested in the development of techniques for fabricating thin-film coils with various planar geometries. One interest in using planar first-order and higher-order gradiometers is to improve the spatial resolution whereby the net magnetic flux from a more distant source is reduced in favor of the closest-lying source. Another motivation is to improve the precision of the coil's field balance, i.e. matching the area-turns ratios of component coils, so that for a given geometry the sensor is insensitive to relatively uniform fields from distant noise sources. Still another motivation is the desire to simplify the process of fabricating such coils, in anticipation of the time a few years hence when arrays of 100 or more detection coils will be used to measure simultaneously the field over the entire scalp.

CRYOSQUID

During the past decade there has been continuing interest[10-13] in developing a closed-cycle refrigerator for cooling SQUID-based magnetometers for a variety of applications, including neuromagnetometry. Such a device is desirable for several reasons, including: reduction of operating costs; possible use in a remote, untended environment; the unavailabilty of liquid helium in certain circumstances; safety; and the convenience of not having to transfer helium every two or three days.

The system we describe is intended to be used for neuromagnetometry both by itself and in conjunction with other neuromagnetometers. Toward that end we established several design goals. The system should be able to operate at or near current dewar-based noise levels, which means a noise level of approximately 20 $fT/Hz^{1/2}$. No cryogenic liquids, either He or N_2, should be required for operation. It should operate for at least one month between shut-downs. It should be able to operate over a range of orientations, including nearly up-side down. And it should also be capable of operating either inside or outside a magnetically shielded room. We have succeeded in developing such a system that meets these goals. To acknowledge the method of cooling the sensor we call this device "CryoSQUID".

We have constructed a prototype hybrid system incorporating a Gifford-McMahon (GM) cycle and a Joule-Thomson (JT) refrigerator designed to provide cooling of a standard BTi DC SQUID and a single second-order gradiometer detection coil of conventional design (Fig. 1). This

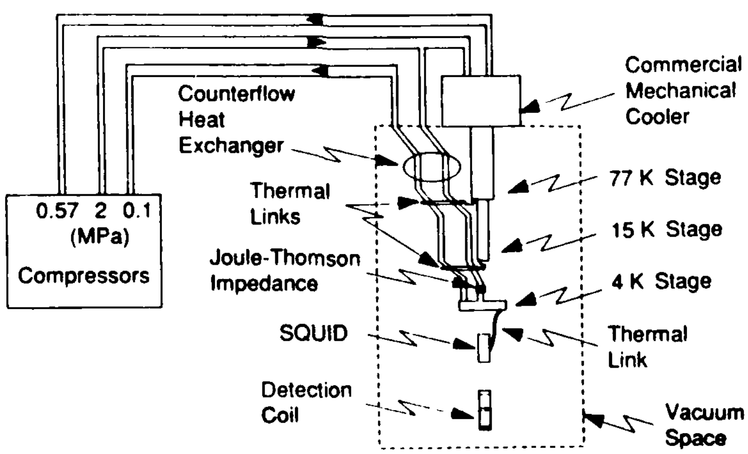


Fig. 1. Schematic for gas flow lines and thermal links in the CryoSQUID system.

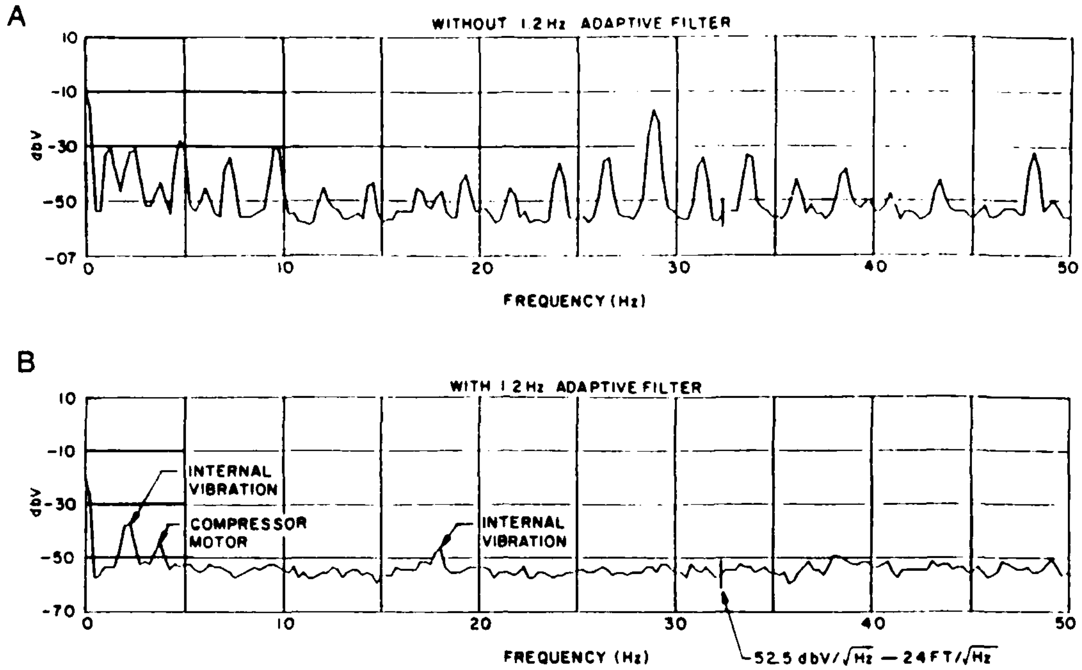


Fig. 2. Noise spectral density of CryoSQUID when the magnetic sensor is placed in a magnetic shield. A) System operating without adaptive filter. B) System operating with the adaptive filter that subtracts a sample of the previously measured noise from the record.

system has operated successfully at equivalent noise levels of 25 $fT/Hz^{1/2}$ over a bandwidth of 1-500 Hz, with minor exceptions where noise peaks are found at a few undamped internal vibrational frequencies of the system (Fig 2). The system is designed around an off-the-shelf, two-stage GM cooler having approximately one watt of cooling at 15 K. To this is added a helium gas JT loop and a suspension system that is designed to minimize the vibration transmitted to the SQUID and detection coil from the GM stage. The JT is run with an input pressure of 2 MPa and a return pressure of 100 kPa. The GM uses the same input pressure and a return pressure of 570 kPa. Two standard helium compressors provide the gas supplies. The current prototype weighs approximately 23 kg and has a cool-down time from room temperature of approximately 34 hours.

The use of a JT cycle for the low temperature stage eliminates the noise problems associated with the temperature fluctuations of an all mechanical design.[12] It also provides the lower temperature (less than 5.5 K) required for low-noise operation of the BTi DC SQUID. Most of the tests performed so far have been done while running the JT loop in an open cycle from a helium storage bottle.

While the suspension system reduces the vibration experienced by the SQUID and pickup coil it does not eliminate it. This is evident in Fig. 2A where at 4.8 Hz the equivalent noise is 550 $fT/Hz^{1/2}$. These data were taken while running the system in a mu-metal can to shield the system from ambient field noise. Results outside of the mu-metal give similar results for the vibrational noise. Because of the stable characteristics of the vibrational noise most of it can be effectively eliminated by the use of a computer-based adaptive filter as shown in Fig. 2B. Here the noise characteristics were monitored and averaged by computer with a reference synchronized to the GM basic period. When the resulting averaged noise is subtracted from the time series data, a relatively flat spectrum is obtained.

We are in the process of building an improved version of this device. This will reduce the weight of the dewar and contents to less than 15 kg, and it should eliminate unwanted internal vibrations. Furthermore, we estimate it should reduce the cool-down time to less than 18 hours. The white noise level of the SQUID system should be reduced to 20 $fT/Hz^{1/2}$ through the elimination of the present excessive rf shielding within the dewar. This version is also being designed to operate on a gantry that will provide the capability for inverting the dewar. This is an important feature, because CryoSQUID can then be used to monitor activity at the side of the upright head, or activity low on the posterior region of the head, near the primary visual cortex of the brain.

MAGNETIC SHIELDING

Hospitals and clinics are notoriously noisy from the standpoint of electromagnetic radiation, low-frequency magnetic noise, and noise at intermediate frequencies contributed by machinery and power lines. Use of a detection coil with the geometry of a second- or higher-order gradiometer overcomes much of this noise in a normal laboratory setting; but greater reliability in the quality of noise reduction may be required for routine clinical applications. A special problem with unshielded measurements occurs at low frequencies, below a few hertz, where environmental noise generally shows an inverse power-law dependence on frequency and rises above the intrinsic sensor noise. This is a spectral region of particular interest for clinical measurements, since signals associated with higher levels of brain function may lie there.[14] Magnetic shielding offers a solution to this problem, as first demonstrated by Cohen *et al.*[15] when recording magnetocardiograms in a multi-layer chamber at the Francis Bitter National Magnet Laboratory at M.I.T. While this shield was constructed with a shape approximating that of a sphere, recently constructed shields[15-17] show that adequate performance can be obtained with a cubic or rectangular shape. The advantage of such a geometry is availability of greater usable interior space. Eddy-current shielding has also been used to advantage in rooms fabricated from thick aluminum plate,[18] but the effectiveness of this shielding is markedly reduced below a few hertz frequency.

We report here the performance of a particularly large magnetically shielded room (MSR) recently installed in the Neuromagnetism Laboratory at New York University.[19] Similar MSRs have been erected during the past half-year at three other institutions. The interior has floor dimensions of 3 m × 4 m and a height of 2.4 m. Such a generous space is important for patient comfort and safety, especially when the patient may need constant attention by a physician. This room consists of an inner shield of mu-metal™ mounted on 8-mm thick aluminum plate that serves as an eddy-current magnetic and radio-frequency shield. The montage is supported by a stiff aluminum framework of 15-cm thickness, and the outer surface of the framework is covered by a second mu-metal shield. The ceiling of the MSR supports aluminum railings which may be used to suspend one or two gantries for holding the dewar containing the magnetic sensors. A single door when closed provides magnetic continuity for the two layers of magnetic shielding. Figure 3 shows the front of this room before cosmetic panels were attached to its surfaces. Access ports are provided for air circulation, optical fibers, and filtered electrical leads.

The attenuation provided by this MSR was evaluated by R.T. Johnson and J.R. Marsden by placing an electromagnet about 5 m or more from the wall of the room with its axis vertical. A SQUID magnetometer positioned at the center of the room detected the interior field strength when fields of various frequencies were applied. Figure 4 illustrates the deduced attenuation over the frequency range of primary interest. After a 60-Hz shaking field was applied, the steady earth's field is attenuated by a factor of 10^3. However, shielding is less effective for very low-frequency fields, being only about 30 db. The effect of eddy-current shielding becomes apparent above about 10^{-1} Hz, and the attenuation rises to a value of about 10^4 at 10^2 Hz with a tendancy toward saturation at higher frequencies. The shielding is not quite as effective as for MSRs having three[17] or more[15] separated layers of magnetic shielding, but the present room is considerably less expensive and has greater interior space.

The improvement in noise level that the MSR provides can be appreciated by examining the spectrum in Fig. 5. This shows the field spectral density measured by J. Shang and B. Schwartz of a SQUID sensor placed near the center of the room and oriented vertically. This sensor, which is one of five mounted on a probe suspended in a single dewar,[8] has a second-order gradiometer as its detection coil, with a baseline between adjacent coils of 4 cm. In the absence of the MSR the low-frequency ambient noise first became apparent over the intrinsic sensor noise at a frequency of 25 Hz. With the MSR it becomes apparent at the considerably lower frequency of about 1 Hz.[20] This substantial improvement is considered very satisfactory although not state-of-the-art. The noise level can be further reduced by improving the area-turns balance of the detection coils, which are known to be out of balance by a few parts in 10^4. This dewar contains three other SQUID sensors in

Fig. 3.

Magnetically shielded room at the Neuromagnetism Laboratory of the Departments of Physics and Psychology at New York University.

addition to the five used to detect neuromagnetic fields. These three have magnetometer detection coils that are oriented to monitor three orthogonal components of the ambient field, so that their outputs can be properly scaled and subtracted from the signal to further improve the noise level. We are now in the process of evaluating how effective this procedure is to further enhance sensitivity within the MSR.

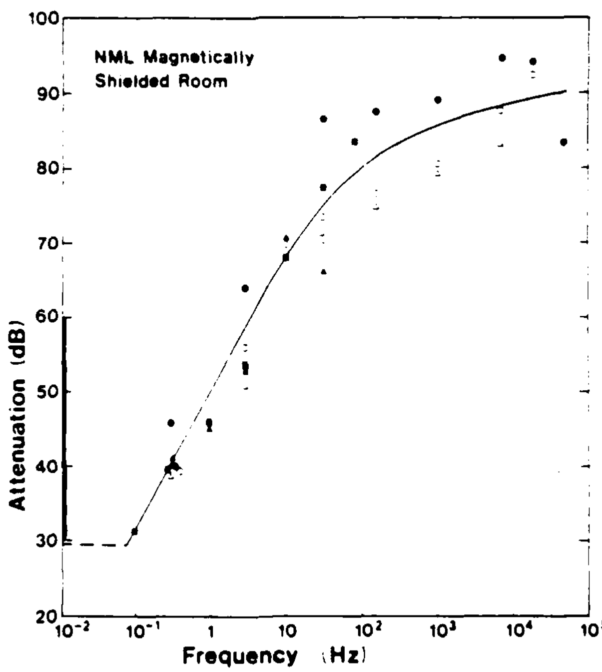


Fig. 4. Attenuation at various frequencies of the MSR shown in Figure 3. Different symbols indicate measurements with the field coil at positions in front or to the side of the room, and at different distances from it.

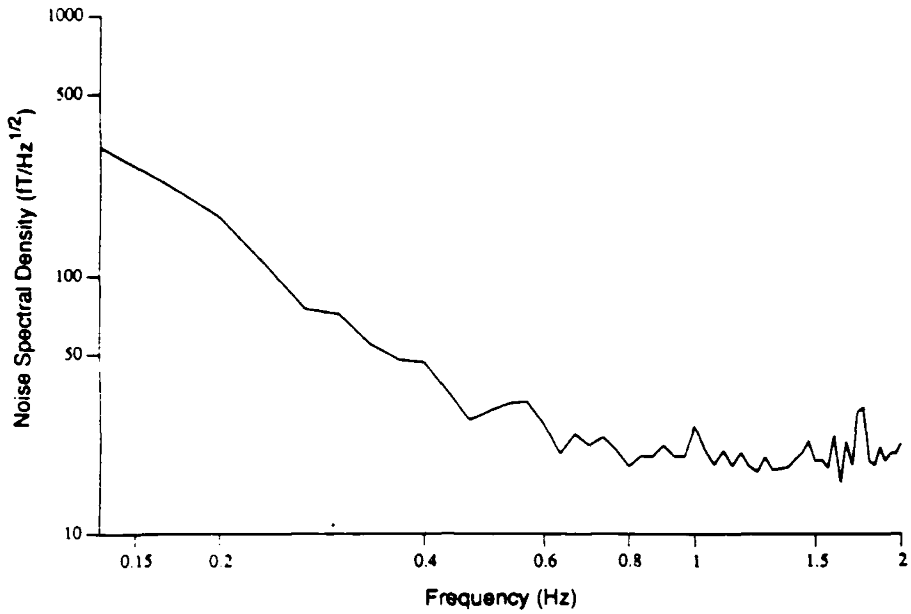


Fig. 5. Field noise indicated by a SQUID sensor placed near the center of the MSR. The high-frequency noise is sensor system noise, and corresponds to a field spectral density of about 20 $fT/Hz^{1/2}$. Environmental noise becomes apparent below about 1 Hz.

DEWAR GANTRY

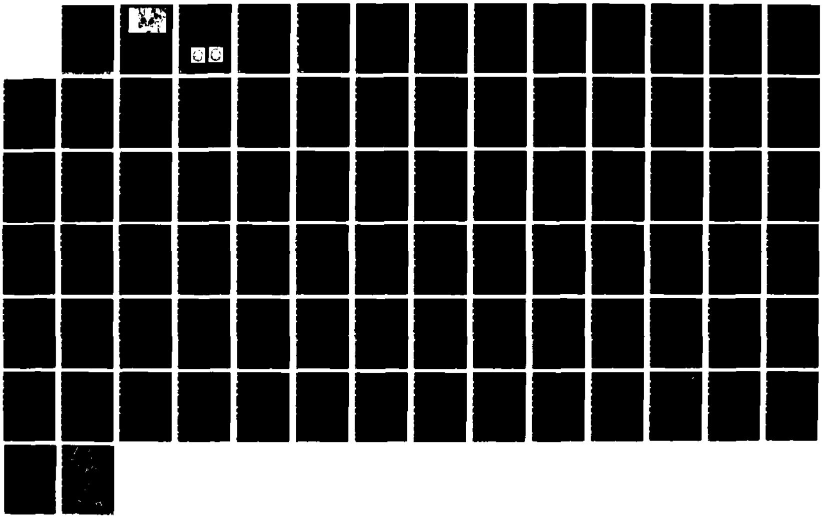
To determine where neural sources lie within the head it is necessary to measure accurately the position and orientation of each field sensor with respect to landmarks on the patient's scalp. Traditionally the dewar was placed in the desired location, and the position of its tail with respect to convenient landmarks was measured across the scalp. But this procedure has inherent inaccuracies, due principally to the irregular shape of the head. One advance was to align the patient's head within a reference framework, and to move the dewar accurately with respect to this framework.[3,8] An example of such a setup is shown in Fig. 3, where the dewar is held in a carriage that moves so that the dewar's axis always points toward the center of the patient's head. This has an advantage when the head is modeled as a sphere for computing source locations, for the field component provided by the sensors is exactly the radial component. Another procedure is to use a computer-controlled mechanized gantry that moves the dewar to a pre-determined position and orientation in space.[20] In Fig. 6 a different approach is shown, where two independent gantries permit the operator to move each dewar by hand to the desired location. Independent movement is provided along two orthogonal horizontal directions and the vertical direction, with rotation allowed about the vertical axis and the horizontal axis where the gantry supports the dewar. Friction holds the dewar in place when the operator releases it, and a secure lock is provided by compressed gas brakes that secure all these degrees of freedom. The dewar can also be rotated about its own axis if it is desired that the individual sensors within the dewar be placed in a particular orientation.

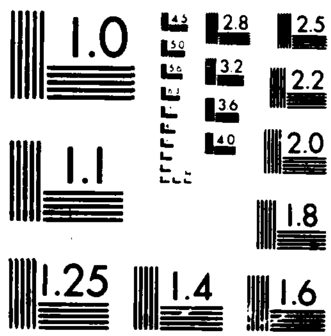
PROBE POSITION INDICATOR

The gantry just described is augmented by a magnetic system that determines the position of each sensor with respect to the patient's head. This is called the "Probe Position Indicator". It consists of a transmitter mounted on the bottom of the main section of each dewar, four receivers, an electronic controller, and computer software to interpret the output. Each transmitter and receiver contains three orthogonally oriented coils. An ac current at about 10 kHz is sent through each of the transmitter coils in turn, and the three voltages induced in the coils of each receiver are measured. The position and orientation of a given receiver is computed from the nine amplitudes of the voltages it provides.

AD-A193 392 COGNITIVE AND NEURAL BASES OF SKILLED PERFORMANCE(U) 2/2
NEW YORK UNIV N Y L KAUFMAN 04 OCT 87 AFOSR-TR-87-1710
F49620-86-C-0131

UNCLASSIFIED F/G 5/8 NL




1.0
2.8
2.5
3.2
2.2
3.6
4.0
2.0
1.1
1.8
1.25
1.4
1.6

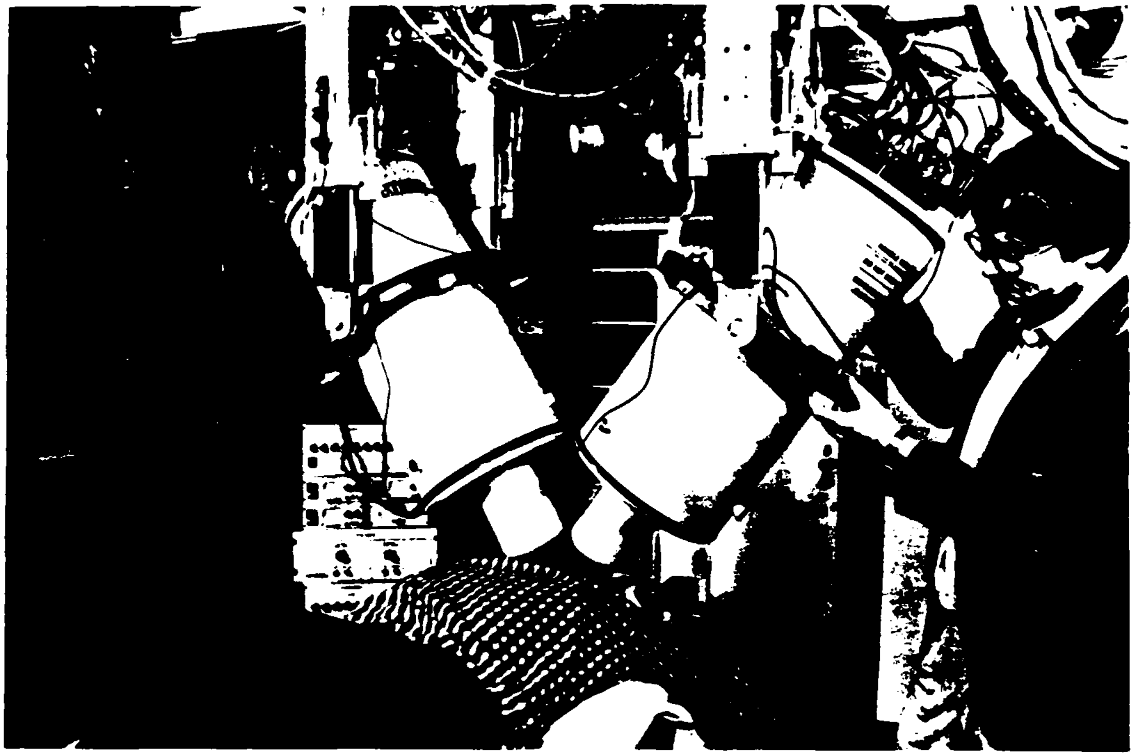


Fig. 6. A pair of dewars, each containing a probe with 7 SQUID sensors, supported by gantries over a subject in the MSR at the Center for Neuromagnetism of the NYU Medical Center.

When recording neuromagnetic data, a patient wears a headband similar to an athlete's sweatband, and from one to three receivers are attached to the band. The fourth receiver, which has a short pointer attached, is used by the operator to point to landmarks on the scalp, so that the positions of these receivers are established with respect to the landmarks. It is necessary to have only one receiver on the headband if there is assurance that it will not move or tip from its original orientation during measurements. However, it is more reliable to use two or three receivers, since inter-comparison of their independent measurements provides an indication of how stable the arrangement is. Between neuromagnetic recordings the operator can determine the position of the the sensors with respect to the subject's head whenever desired.

Measurements carried out by J. Shang have shown that the accuracy of this Probe Position Indicator is generally better than ±2 mm when operating in the shielded room depicted in Fig. 6, provided the tails of the dewars are close to the scalp. Accuracy decreases if the dewars are moved further than about 20 cm away. In practice, with this system, the accuracy in determining sensor positions is limited by how reproducibly the landmarks can be identified by the operator and not by the accuracy of the Probe Position Indicator itself.

CLINICAL APPLICATIONS

Ongoing clinical research at the University of California at Los Angeles[21,22] and at the Istituto di Neurochirurgia of the Università degli Studi di Roma[3,23] have revealed a number of significant applications of magnetic localization in studies of tumors and epilepsy. The presence of a tumor, even those too small to be resolved by x-ray CT or magnetic resonance images, may produce abnormal magnetic activity that can be localized.[3] However, most of the investigations reported to date have concentrated on epilepsy, which is a disorder characterized by the abnormal electrical discharge of neurons within the brain that result ultimately in behavioral seizures. Localized discharges, or focal epilepsy, is the most common, and it is estimated that some 800,000 individuals in the United State alone have such a disorder. Most can be treated effectively with medication; however, this is insufficient for as many as 360,000 of them. For perhaps 15% of the latter group, surgical removal of the epileptic brain tissue may be considered. Obtaining a precise localization of the focal region is of paramount importance in deciding whether to intervene surgically.

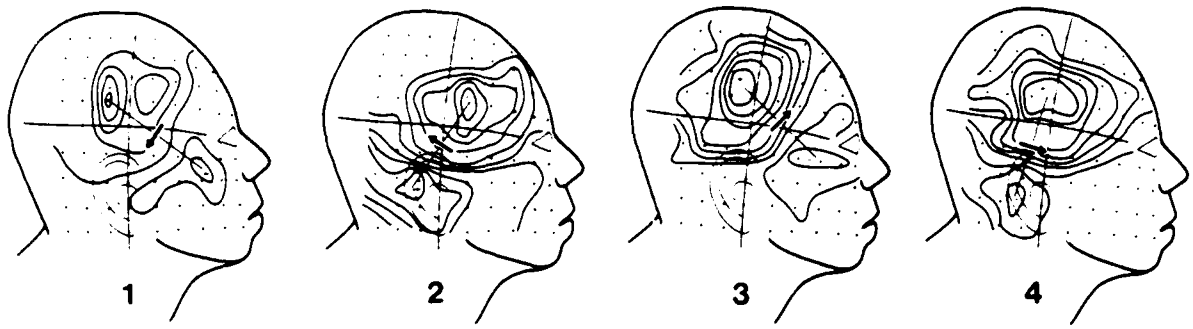


Fig. 7. Isofield contours for a four-wave sequence of interictal epileptic activity. The arrows show the position and direction of the underlying current sources. From Ref. 21.

The discharges seen in patients have a variety of forms. Some can be localized to a single position, while others have a more complex origin. Figure 7 shows the sequence of field patterns from a complex series of discharges in a young male patient, as reported by Barth *et al.*[21] Two different sources generate biphasic discharges in an interleaved sequence. The first map indicates a discharge in the right anterior temporal lobe with the current directed downward and posteriorly. The second, recorded 16 milliseconds after the first, indicates a discharge from a different source lying 1 cm behind and below the first. The third discharge is from the first source with the current reversed, and the last discharge is identified as originating from the second source, with reversed current. Viewed magnetically, it is often possible to unravel such complex spatial and temporal discharge patterns that cannot be interpreted from the electroencephalogram alone.

In this particular example, the intracranial locations of the first and second sources are reasonable from the brain's anatomy and observed pathology. Both sources lie at the edges of a large region of scar tissue within the right temporal lobe. Figure 8 shows x-ray CT scans on which the computed source positions are shown by crosses. At the borders of scarring, functioning neurons are often disturbed, thus producing intermittent epileptic discharges. Accumulating evidence such as this, together with a large body of data obtained in studies of normal brain activity, provides verification of existing procedures for neuromagnetic localization.

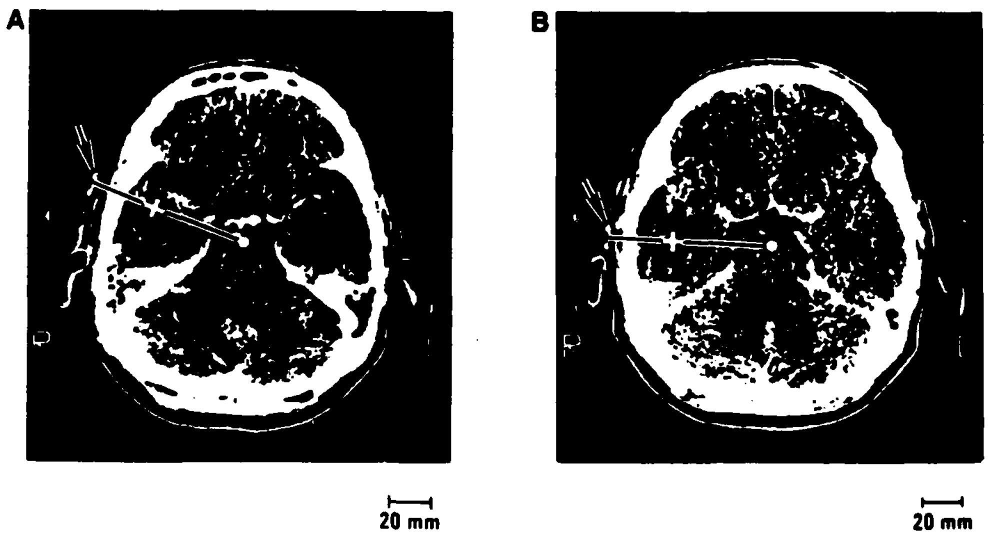


Fig. 8. Computerized tomography scans showing deduced locations for the first and second sources that give rise to the epileptic activity in Fig. 7, from Ref. 21.

ACKNOWLEDGEMENTS

We express our sincere appreciation to L. Kaufman, D.B. Crum, R.J. Ilmoniemi, R.T. Johnson, J.R. Marsden, B. Schwartz, and J. Shang for their help and informative discussions. Research at New York University is supported in part by Air Force Office of Scientific Research Contracts AFOSR-84-0313 and F49620-86-C-0131.

REFERENCES

1. D. Cohen, Magnetoencephalography: evidence of magnetic fields produced by alpha rhythm currents, Science 161:784 (1968).
2. D. Cohen, Magnetoencephalography: detection of the brain's electrical activity with a superconducting magnetometer, Science 175:664 (1972).
3. G. L. Romani and L. Narici, Principles and clinical validity of the biomagnetic method, Med. Prog. thru. Tech. 11:123 (1986).
4. R. Hari and R. J. Ilmoniemi, Cerebral magnetic fields, CRC Critical Rev. in Biomed. Eng. 14:93 (1986).
5. S. J. Williamson and L. Kaufman, Analysis of neuromagnetic signals, in: "Handbook of Electroencephalography and Clinical Neurophysiology," A. Gevins and A. Rémond, eds., Elsevier, Amsterdam (1987), Chapter 14.
6. S. J. Williamson, G. L. Romani, L. Kaufman, and I. Modena, "Biomagnetism: An Interdisciplinary Approach," Plenum Press, New York (1983).
7. R. Ilmoniemi, R. Hari, and K. Reinikainen, A four-channel SQUID magnetometer for brain research, Electroenceph. clin. Neurophysiol. 58:467 (1984).
8. S. J. Williamson, M. Pelizzone, Y. Okada, L. Kaufman, D. B. Crum, and J. R. Marsden, Magnetoencephalography with an array of SQUID sensors, in: "Proceedings of the 10th International Cryogenic Conference - ICEC10", Butterworth, London (1984), p. 339.
9. G. L. Romani, R. Leoni, and C. Salustri, Multichannel instrumentation for biomagnetism, in: "SQUID '85: Superconducting Quantum Interference Devices and their Applications," H. D. Hahlbohm and H. Lübbig, eds., Walter de Gruyter, Berlin (1985), p. 919.
10. J. E. Zimmerman and R. Radebaugh, Operation of a SQUID in a very low-power cryocooler, in: "NBS Special Publication 508," National Bureau of Standards (1978).
11. J. E. Cox and S. A. Wolf, Magnetic and vibrational characteristics of a close cycle refrigerator, in: "NBS Special Publication 508," National Bureau of Standards (1978).
12. D. B. Sullivan, J. E. Zimmerman and J. T. Ives, Operation of a Practical SQUID Gradiometer in a Low-Power Stirling Cryocooler, in: "NBS Special Publication 607," National Bureau of Standards (1981).
13. E. Tward and R. Sarwinski, A closed cycle cascade Joule Thomson refrigerator for cooling Josephson junction devices, in: "NBS Special Publication 698," National Bureau of Standards (1985).
14. Y. Okada, L. Kaufman, and S. J. Williamson, Hippocampal formation as a source of endogeneous slow potentials, Electroenceph. clin. Neurophysiol. 55:417 (1982).
15. A. Mager, The Berlin magnetically shielded room (BMSR), Section A: Design and construction, in: "Biomagnetism," S.N. Erné, H. D. Hahlbohm, and H. Lübbig, eds., Walter de Gruyter, Berlin (1981), p. 51.
16. M. Ibuka, H. Hosomatsu, and S. Naito, A SQUID magnetometer using a niobium thin-film microbridge, IEEE Trans. Instrum. and Meas. IM-30:251 (1981).
17. V.O. Kelhä, J. M. Pukki, R. S. Peltonen, A. J. Penttinen, R. J. Ilmoniemi, and J. J. Heino, Design, construction, and performance of a large-volume magnetic shield, IEEE Trans. Magn. MAG-18:260 (1982).
18. J. E. Zimmerman, SQUID instruments and shielding for low-level magnetic measurements, J. Appl. Phys. 48:702 (1977).
19. Fabricated by Vacuumschmelze GmbH, Hanau, Federal Republic of Germany.
20. J. Vrba, M. Burbank, H. Ensing, A. Fife, E. Heijster, C. Marshall, J. McCubbin, D. McKenzie, M. Tillotson, K. Watkinson, H. Weinberg, and P. Brickett, Integrated biomagnetic robotic system, in: "Biomagnetism: Applications and Theory," H. Weinberg, G. Stroink, and T. Katila, eds., Pergamon Press, New York (1984), p. 52.

21. D. Barth, W. Sutherling, J. Engle, Jr., and J. Beatty, Neuromagnetic evidence of spatially distributed sources underlying epileptiform spikes in the human brain, Science 223:293 (1984).
22. J. Beatty, D.S. Barth, and W. Sutherling, Magnetically localizing the sources of epileptic discharges within the human brain, Naval Res. Rev. 2:20 (1984).
23. G.B. Ricci, Clinical magnetoencephalography, Nuovo Cimento 2D:517 (1983).

AFOSR-TR- 87-1710

METHODS AND INSTRUMENTATION FOR BIOMAGNETISM

S.J. Williamson, S.E. Robinson,* and L. Kaufman

Neuromagnetism Laboratory, Departments of Physics and Psychology
New York University, 4 Washington Place, New York, NY 10003, U.S.A.
*Biomagnetic Technologies, inc., 4174 Sorrento Valley Boulevard
San Diego, CA 91630, U.S.A.

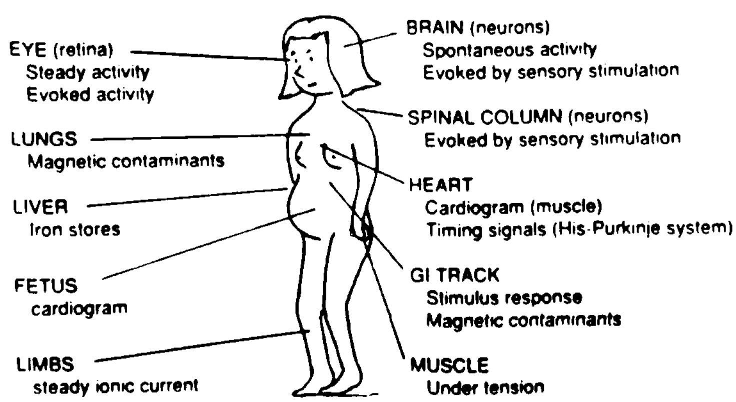
INTRODUCTION

Biomagnetism is the study of biological activity through analysis of the magnetic fields such activity produces. This paper is a brief explanation of the principles of biomagnetism, focussing on the instrumentation that make such studies possible and how these measurements are used to learn about the underlying biological structures and events that can be deduced from them. This is not meant to be a comprehensive review, and some important areas of study will not even be mentioned. Readers who wish a more complete coverage may consult a textbook which provides an extensive introduction to this broad topic (Williamson et al., 1983), the proceedings of the last international conference on biomagnetism (Weinberg et al., 1984), and a general review (Williamson and Kaufman, 1981).

BIOMAGNETIC FIELDS

Many organs of the human body produce magnetic fields, as depicted in Fig. 1. There are three classes of sources: magnetic materials, the magnetic susceptibility of tissue, and ionic electrical currents. The first biomagnetic field to be observed from magnetic materials was associated with particles lodged in the lungs, as well as other organs in the thorax (Cohen, 1975). More difficult to observe is the effect of tissue susceptibility, because its value is close to that of water, the body's major constituent, and its value is quite low (Farrell et al., 1978). Nevertheless for patients with substantial iron overloads in the liver, measurements of that organ's susceptibility *in vivo* provide a clinically important measure of the concentration of iron (Brittenham et al., 1982). The class of fields that has attracted the most interest are those arising from electrical currents in the body. The strongest field is associated with the strongest current, that of the heart muscle (Cohen, 1970). The earliest biomagnetic studies focused on mapping at various places across the chest the time-course of cardiac activity, called the *magnetocardiogram* (MCG). A related subject of prime interest is the conduction system of the heart, including the His bundle and Purkinje system that carry excitations from the pacemaker to the ventricles. While these rapidly-moving signals are difficult to observe, the clinical importance of developing a noninvasive technique to monitor the conduction system has encouraged intensive research (Erné, 1985). The greatest emphasis has been concentrated on much weaker signals from neural activity within the brain, or *magnetoencephalogram* (MEG). Studies of spontaneous and sensory-related brain activity have demonstrated

Fig. 1. Representative magnetic fields of the human body.



their importance for both basic and clinical research (For reviews see: Hari and Ilmoniemi, 1986; Romani and Narici, 1986; Williamson and Kaufman, 1987). Even the very weak fields associated with brainstem activity have been detected (Erné et al., 1987). Recently, magnetic fields have also been observed in the vicinity of the spinal cord in humans (Mizutani and Kuriki, 1986).

These kinds of studies have stimulated interest in better understanding the underlying physiology that gives rise to the fields. Thus Wikswo et al. (1980) studied the field associated with the action potential of an isolated nerve axon *in vitro* and demonstrated that the observed field is due to intracellular currents. Simultaneous measurements of both transmembrane potential and the magnetic field near a nerve have validated the underlying theory and provide accurate measurements for the conductivity of the intracellular medium (Roth and Wikswo, 1985). In a similar spirit, research has begun on isolated brain tissue to gain understanding of the underlying mechanisms when populations of neurons are active (Okada and Nicholson, 1987; Tesche et al., 1987).

INSTRUMENTATION

While studies of these types have secured for the biomagnetic approach an accepted place in a variety of specialized disciplines, work continues toward developing improved measuring techniques. All biomagnetic fields are extremely weak, the strongest being about 10^{-6} of the earth's steady field of 70×10^{-6} tesla (or 70 μT). Thus the QRS peak of the cardiac field is typically 25×10^{-12} T (or 25 pT), the much weaker alpha rhythm of the brain is about 1×10^{-12} T (or 1 pT) and sensory-evoked fields are about 100×10^{-15} T (or 100 fT). In virtually all cases the investigator must cope with two problems: the weakness of the signal and the strength of competing magnetic noise in the environment. Here we discuss only the basic concepts of how to deal with these problems, for details can be found in a recent review (Romani et al., 1982a).

SQUID Sensors

A cryogenic instrument known as a *superconducting quantum interference device* (SQUID) is used for the most sensitive biomagnetic studies. It is conventionally maintained in a bath of liquid helium at a temperature of 4.2 K (or -269 C), isolated from the outside by a vacuum-insulated contained known as a *dewar* whose external surfaces are at room temperature (Fig. 2). Two components of this system merit special attention: one is the *SQUID* itself and the other is a *detection coil* that is placed as close as possible to the field source. The detection coil is part of a closed-loop superconducting circuit, called a *flux transformer*, with the leads of the coil passing upward to enter the SQUID enclosure where they form a *signal coil*. One property of a closed superconducting loop is that if a magnetic field is applied anywhere within the loop, the superconducting electrons flow through the wire so that their current produces a field that maintains the net magnetic flux in the loop (product of field and area) invariant. Consequently, if a magnetic field passes through loops of the detection coil, current passes around the entire circuit, and the portion flowing in the signal coil imposes a field on the SQUID. Room temperature electronics monitor the response of the SQUID and provide a voltage that is proportional to the magnetic flux in the detection coil. The principal advantage of this arrangement is that the detection coil can be wound with a geometry best meeting the measurement

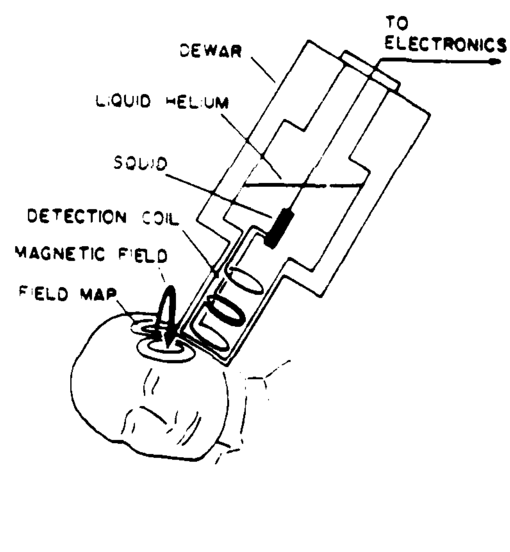
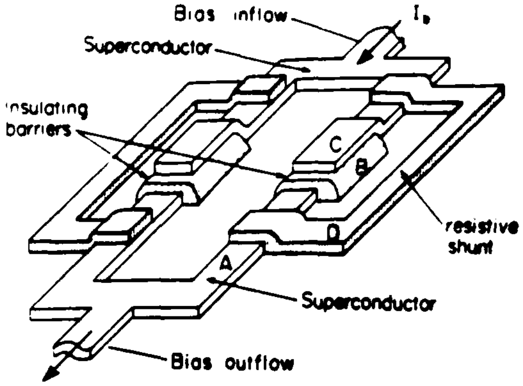


Fig. 2. Elements of a SQUID system for monitoring the magnetic field of a subject's brain.

Fig. 3. A thin-film dc SQUID, with A and C being superconducting layers and B a thin insulator. Each junction has a parallel resistive shunt (D) for technical reasons, to avoid a hysteretic response when a magnetic field is applied.



requirements. For instance, Fig. 2 shows a coil having the geometry of a *second-order gradiometer*, namely, three equally-spaced coaxial coils with the center wound in the opposite sense and having twice as many turns as the end coils. This arrangement discriminates against fields from distant noise sources that are uniform in space, yet it retains sensitivity to fields from local sources that are appreciable only in the lowest coil (*pickup coil*). Thus the second-order gradiometer improves the signal-to-noise ratio. In less-noisy environments, such as a rural site or inside a magnetically shielded room, a first-order gradiometer may suffice. This has only two separated coils, wound in the opposite sense. In all gradiometers, the quality of noise rejection depends on how accurately the areas enclosed by each loop are made equal to each other. It is common to attach small pieces of superconducting foil at the appropriate positions to improve the area "balance" to 1 part in 10^4 or even 1 part in 10^5. Detection coils of large diameter enhance sensitivity because they couple more signal energy. Typical diameters are 2 and 4 cm for studies of the brain and heart respectively.

With the advance of technology, thin-film techniques are becoming popular for fabricating SQUIDs. Figure 3 gives a simplified illustration of what is known as a *dc SQUID*. To operate this device a dc current that is fed into the superconducting film (C) at one end, divides and passes through two parallel arms, and recombines in the superconducting film (A) at the other end. Each arm is interrupted by a *Josephson junction*, which is simply a thin insulator (B) that breaks the superconducting circuit. This is named for Brian Josephson who developed the theory for the action of such an insulating juction between two superconductors. At low dc bias current, electrons can "tunnel" through the junctions without exhibiting resistance, but when the current is increased to an appropriate level both junctions exhibit a (common) voltage. This voltage is predicted by Josephson's equations and the electrical behavior of the circuit.

If a signal coil mounted above the area between the two arms applies a magnetic field, the voltage across the junctions varies periodically with increasing field. This arises from the fact that the field gives electrons passing along one arm a different momentum than electrons passing along the other, so there is an interference effect varying with field when the quantum-mechanical wave functions describing the two currents rejoin. The condition for periodicity is fixed by the value of the elementary *flux quantum* $\phi_0 = h/2e = 2 \times 10^{-15}$ tesla-meter^2, which is the smallest non-zero amount of flux that can exist within a closed loop of superconductor. Since the area within the arms of the SQUID is small, and the flux quantum itself is such a small value, counting voltage oscillations provides a senstivity measure of how much the applied field has changed.

This device can be made to respond linearly with applied field, and the sensitivity can be enhanced by a factor of 10^5 or more by adding a feedback loop. With a second coil mounted over the area between the arms and appropriate electronic circuits to monitor the voltage across the junctions, a current can be fed to this coil so that its field just cancels the field of the input coil. With the SQUID serving as a null detector in this way, the voltage provided by the feedback current passing through a resistor is strictly proportional to the current of the signal coil, and hence to the biomagnetic field in the detection coil. The method provides excellent linearity in response, with a wide dynamic range. Other refinements are added to improve sensitivity, such as applying an ac rather than true dc bias, but it is not appropriate to go into such details here. Some commercial SQUID devices with detection coils of 2-cm diameter exhibit a sensitivity of about 20 fT within a 1 Hz bandwidth, and with careful optimization a sensitivity of 5 fT has been achieved. Emphasis has also been placed in developing multiple-sensor systems so that the process of mapping a field pattern over a portion of the body to determine the underlying sources is greatly shortened (Ilmoniemi et al., 1984; Williamson et al., 1984; Romani et al., 1985; Knuutila et al., 1987). Since thin-film fabrication techniques offer many advantages for SQUIDs, there is interest in making detection coils in the same way. Indeed, with sufficient sensitivity, the use of higher-order planar gradiometers (as contrasted with the axial gradiometer shown in Fig. 2) may have advantages in localizing sources of biomagnetic signals.

Magnetic Shielding

SQUID systems using a second-order gradiometer for the detection coil are capable of a wide range of useful measurements in an unshielded setting, including laboratories and clinics. However, ambient noise increases dramatically at low frequencies in noisy locations, and there may be additional noise at discrete frequencies from nearby machinery. High levels of radio frequency noise, as from communication systems, may also interfere with the operation of the SQUID. A room constructed with magnetic and radio-frequency shielding is one effective way to minimize these problems. The first rooms built for biomagnetic applications had four or more widely-spaced, concentric shells of high-permeability material (e.g., Mager, 1981) and are very effective. However, this requires a large space for installation, and the space inside is small, having a characteristic dimension of 2 m. Other magnetically shielded rooms have fewer shells, provide more working space, and yield acceptable shielding for most purposes (Kelhä et al., 1982; Buchanan et al., 1987).

SQUID Refrigeration

Everyone who uses a SQUID system recognizes that liquid helium as a coolant is a nuisance and considerable expense. There would be considerable advantage in using a refrigerator to cool and maintain the low temperature portions of the dewar. A SQUID requires very little refrigeration capacity, and the capacity to cool the electrical leads and insulating vacuum section is modest. The principal challenge is to limit the magnetic noise and vibration that such a system imposes on measurements. This was recognized by Zimmerman and Radebaugh (1978) who developed a successful closed-cycle cryogenic refrigerator, or *cryocooler*, based on a Sterling cycle. This has a long, motor-driven displacer that moves within a close-fitting sleeve to admit helium gas under pressure and subsequently achieve cooling by expansion of the gas as the displacer is withdrawn. It requires only 50 W of input power to produce temperatures of about 7 K, sufficient to operate a niobium rf SQUID that detected the magnetic field of the human heart.

Recently a different type of device has been developed with a noise level that is sufficient for measuring the magnetic field of the brain (Buchanan et al., 1987). It depends on both a commercial Gifford-McMahon refrigerator and a specially designed Joule-Thomson refrigerator, where high-pressure helium gas expands and cools as it passes through a small hole into a low-pressure region. The former cools the dewar from room temperature and maintains a 15 K stage, and the latter is suspended from this cold stage and produces a stable temperature of about 4 K. Figure 4 shows the arrangement of gas lines and thermal links, most of the latter fabricated of fine copper wire. Since there are no mechanical links to the dewar, only fine tubes for conducting helium gas, this device can be easily rotated to operated in nearly any orientation, including horizontally and almost upside down. This makes it especially attractive for measurements about the head. We call this device "CryoSQUID". Movement of the G-M displacer produces magnetic noise, but only at well-defined frequencies. In practice, the noise is so stable that obtaining an average of its time series for 15 seconds or so with a personal computer is sufficient to subsequently allow the computer to subtract this noise in real time

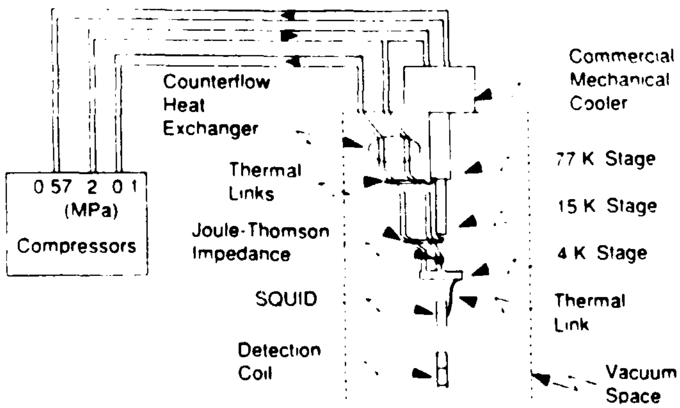


Fig. 4. Schematic for gas flow lines and thermal links in the CryoSQUID system.

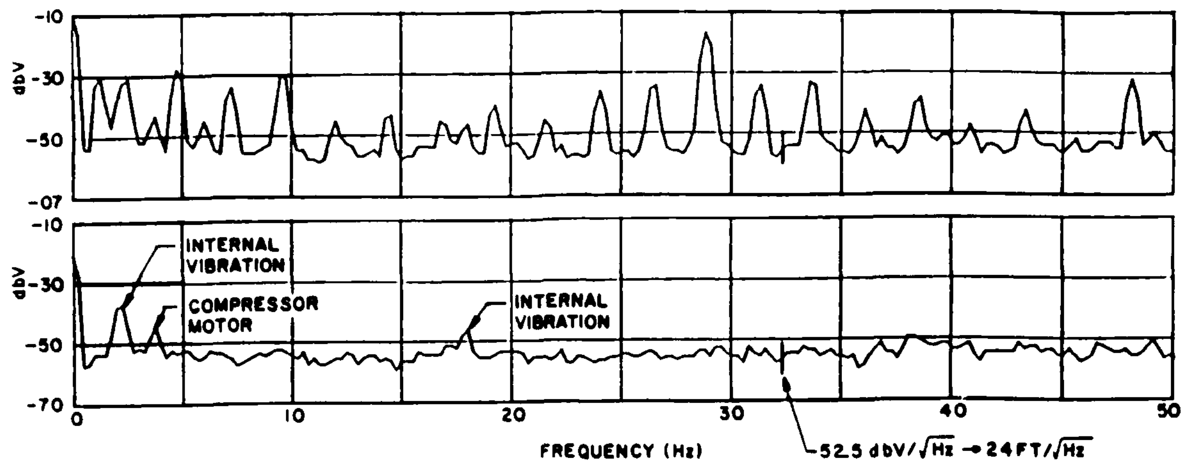


Fig. 5. Noise spectral density of CryoSQUID when the system operates without an adaptive filter (upper panel) and with it (lower panel).

from the incoming data. Figure 5 shows that virtually all of the refrigerator noise is eliminated. With this success it is clear that refrigerator-based SQUID systems are now feasible, elminating the dependence on a steady supply of liquid helium.

PHENOMENA AND METHODS OF STUDY

Magnetic Particles

Magnetic particles exhibit a remanent field once a magnetic field of sufficent strength is applied and then removed. This remanent field is generally permanent, and so the presence of small quantities of magnetic materials can be detected with high-sensitivity sensors. In this way changes in the amount or distribution of particles can be inferred (Cohen, 1975). The technique has been applied to assessment of occupational health in environments where magnetic particles are inhaled (Kalliomäki et al., 1976; Freedman et al., 1980). It has also been exploited to provide evidence that smokers clear particles from their lungs more slowly than non-smokers (Cohen et al., 1979). More recently the study of the remanent field from the lungs of small animals where dust had been intentionally introduced into the lungs has revealed that time-dependent phenomena are caused by cellular activity. The steady decline of the remanent field outside the chest after application of a strong magnetizing field can be attributed to rotation of the particles that have been taken up by cells called "macrophages", which are - so to speak - the garbage trucks of the lung (Gehr et al., 1983a). Corresponding behavior has also been discovered within cells of the liver (Gehr et al., 1983b). Thus, studies of this type are revealing aspects of cell physiology that cannot be obtained noninvasively by other techniques (Nemoto, et al. 1985). Measurements of this kind can be carried out by a device called a fluxgate magnetometer, which requires no cooling.

Ionic Currents

Application of biomagnetic techniques to studies of organs such as the heart and brain are motivated by several factors. One is the fact that a magnetic measurement represents a different kind of spatial weighting of the source currents than electrical measurements across the skin, thus suggesting that different kinds of information will be obtained (Wikswo, 1983). Another is the possibility that the nature of intervening tissue may be less important in influencing the field pattern than the potential pattern across the skin. The overriding advantage of magnetic measurements in our opinion is the possibility of determining more accurately the location of confined regions of activity, when the activity can be modeled as a *current dipole*, namely, a small element of current. Localization provides a means of relating observed activity to specific regions of the body, such as in establishing functional maps of the brain. One example is discovery of a tonotopic organization across the auditory cortex of human subjects (Romani et al., 1982b) or in defining the region of an infarct in a diseased heart (Saarinen et al., 1985; Gonnelli et al., 1985).

Locating a Source

If the source is a current dipole, there will be one region where the field is strongest emerging from the body and another region where it is strongest entering the body. For a body with certain types of symmetry, such as one which can be approximated by a flat surface covering a semi-infinite, uniform conducting region (the *half-space model*), the source lies midway between these field extrema and at a depth that is equal to the distance between the extrema divided by $\sqrt{2}$ (≈ 1.4). Similarly, for a current dipole in a sphere, the depth of the source can be deduced from the ratio of the distance between the extrema to the radius of the sphere (Williamson and Kaufman, 1982). Such simple recipes, which are useful for making first estimates, can be refined by more accurate numerical models describing the appropriate region of the body.

There are cases where the field patterns from two or even three simultaneously active sources have been analyzed to reveal the positions of underlying activity; but in general the problem of dealing with multiple sources is just in its infancy. Much more theoretical and experimental work is needed to deal with the more interesting problems of the time sequence of multiply active areas in the brain or the interplay of Purkinje system and myocardium in activating the heart. In all these cases it should be kept in mind that there is no unique solution for the configuration of electrical sources that can be deduced from electrical potential measurements, or magnetic field measurements, or a combination of the two. Electrically and magnetically "silent" sources exist, in the sense that there are source configurations that produce no skin potential distribution or external fields. Roughly speaking, skin potentials and magnetic fields provide complementary information, and there are cases where the two should be used together (Wood et al., 1985).

FUTURE PROSPECTS

Software Reduction of Noise

Biomagnetic measurements continue to face the problem of environmental magnetic noise when applied to the weakest signals. Most magnetically shielded rooms do not completely shield noise at very low frequencies, say below 1 Hz. With growing popularity in the use of systems with multiple SQUID sensors, it becomes feasible to dedicate one or more of the sensors to monitor the ambient noise as a *reference* for purposes of subtracting a portion of it from the noisy signal. Simple fixed electronic balancing techniques have already been applied with success for unshielded measurements (Williamson et al., 1984). We report here a new computer-based approach that in many practical applications markedly reduces excess low frequency noise. The references consist of three SQUID sensors oriented to monitor three mutually perpendicular components of the ambient field. The technique uses computer-controlled attenuators to adjust the amplitude of each reference that is subtracted from each of the signals so as to remove the correlated portion of the noise. Figure 7 gives an example of the kind of improvement that can be obtained within a magnetically shielded room. With such electronic noise cancellation, the noise level is essentially the intrinsic sensor noise from high frequencies down to a frequency below 0.1 Hz. This is sufficient to operate the SQUID sensors in a dc-coupled mode to monitor very low-frequency activity.

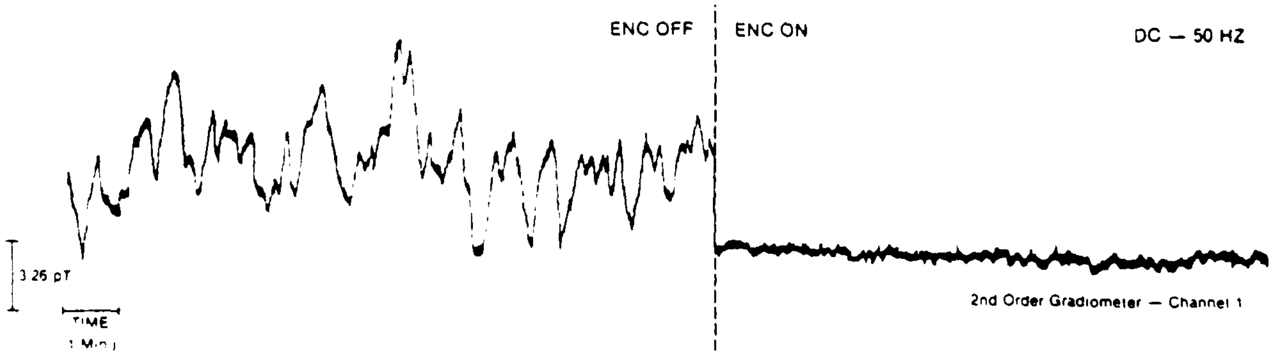


Fig. 7. Time course of the noise in the bandwidth DC-50 Hz observed within the magnetically shielded room at the Center for Neuromagnetism at Bellevue Hospital of the New York University Medical Center, without electronic noise cancellation (left) and with it (right).

High Temperature Superconductors

This past year marked a turning point in the development of superconducting devices with the discovery of superconductivity in the Yttrium-Barium-Copper-Oxygen ceramic system at temperatures as high as 94 K. The race is underway to find ways to make these materials practical for wires, thin films, and SQUIDs. One group recently reported success in making SQUIDs (Zimmerman et al., 1987) that operate up to 80 K. This clearly demonstrates that SQUIDs can be operated at liquid nitrogen temperatures (77 K), but unfortunately the noise levels are several orders of magnitude greater than those of the best SQUIDs operating in liquid helium. Whether high-temperature SQUIDs become useful for biomagnetic applications remains to be seen. Nevertheless, there is reasonable hope that high-temperature detection coils can be fabricated, and this would greatly ease the cryogenic problem. Dewars can be made thinner near the scalp, with a coil operating at 60 or 70 K, and since superconductors are poor conductors of heat the SQUID would not suffer. The prospect for applications of room-temperature superconductors is exciting indeed! When we think of arrays of sensors to measure the field pattern over a large area over the thorax or head, the chief advantage comes from being free from the constraints of a rigid dewar. Then the detection coil positions can be adjusted easily to match the contours of the particular individual. This will enhance signal strengths and permit convenient measurements for children and adults alike.

ACKNOWLEDGEMENTS

We thank D.B. Crum and R.J. Ilmoniemi for helpful discussions. Research at New York University was supported in part by Air Force Office of Scientific Research Grant F49620-85-K-0004 and by Aerospace Medical Research Laboratory Contract SCEE-HER/86-14.

REFEFENCES

Buchanan, D.S., Paulson, D., and Williamson, S.J. (1987). Instrumentation for clinical applications of neuromagnetism. In: Fast, R.W., Ed., Advances in Cryogenic Engineering , Plenum Press, New York.

Brittenham, B.M., Farrell, D.E., Harris, J.W., Feldman, E.S., Danish, E.H., Muir, W.A., Tripp, J.H., and Bellon, E.M. (1982). Magnetic susceptibility measurement of human iron stores. New Eng. J. Med. 307, 1671-1675.

Cohen, D. (1975). Measurements of the magnetic fields produced by the human heart, brain, and lungs. IEEE Trans. Magn. MAG-11, 694-700.

Cohen, D., Edelsack, E., and Zimmerman, J.E. (1970). Magnetocardiograms taken inside a shielded room with a superconducting point-contact magnetometer. Appl. Phys. Lett. 16, 278-280.

Cohen, D., Arai, S.F., and Brain, J.D. (1979). Smoking impairs long-term dust clearance from the lung. Science 204, 514-517.

Erné, S.N., and Fenici, R.R. (1985). The present state of magnetocardiography. In: Collan, H., Berglund, P., and Krusius, M., Eds. Tenth International Cryogenic Engineering Conference, Butterworth, Guildford, pp. 329-338.

Erné, S.N., Scheer, H.-J., Hoke, M., Pantew, C., and Lütkenhöner, B. (1987). Brainstem auditory evoked magnetic fields in response to stimulation with brief tone pulses. Int. J. Neuroscience, in press.

Farrell, D.E., Tripp, J.H., Zanzucchi, P.E., Harris, J.W., Brittenham, G.M., and Muir, W.A. (1980). Magnetic measurement of human iron stores. IEEE Trans. Magn. MAG-16, 818-823.

Freedman, A.P., Robinson, S.E., and Johnston, R.J. (1980). Non-invasive magnetopneumographic estimation of lung dust loads and distribution in bituminous coal workers. J. Occup. Med. 22, 613-618.

Gehr, P., Brain, J.D., Nemoto, I., and Bloom, S.B. (1983a). Behavior of magnetic particles in hamster lungs: Estimates of clearance and cytoplasmic motility. J. Appl. Physiol. 55, 1196-1202.

Gehr, P., Brain, J.D., Bloom, S.B., and Valberg, P.A. (1983b). Magnetic particles in the liver: a probe for intracellular movement. Nature 302, 336-338.

Gonnelli, R., Galeone, P., and Sicuro, M. (1985). A biomagnetic approach to the localization of the necrotic area in inferior myocardial infarction. Med. & Bio. Eng. & Comput. 23 Suppl., Part 1, 59-60.

Hari, R., and Ilmoniemi, R.J. (1986). Cerebral magnetic fields. CRC Critical Rev. in Biomed. Eng. 14, 93-126.

Ilmoniemi, R.J., Hari, R., and Reinikainen, K. (1984). A four-channel SQUID magnetometer for brain research. Electroenceph. clin. Neurophysiol. 58, 467-473.

Kalliomäki, P.-L., Karp, P.J., Katila, T., Mäkipää, Saar, P., and Tossavainen, A. (1976). Magnetic measurements of pulmonary contamination. Scan. J. Work Environ. and Health 4, 232-239.

Kelhä, V., Pukki, J.M., Peltonen, R.S., Penttinen, A.J., Ilmoniemi, R.J., and Heino, J.J. (1982). Design, construction, and performance of a large-volume magnetic shield. IEEE Trans. Magn. MAG-18, 260-270.

Knuutila, J., Ahlfors, S., Ahonen, A., Hällström, Kajola, M., Lounasmaa, O.V., Vilkman, V., and Tesche, C. (1987). A large-area low-noise seven-channel DC SQUID magnetometer for brain research. Rev. Sci. Instrum., in press.

Mager, A. (1982). Magnetisch abgeschirmte Kabine zur Aufnahme kleinster magnetischer und elektrischer Biosignale. Naturwiss. 69, 383-388.

Maniewski, R., Derka, I., Katila, T., Ryhänen, T., and Seppänen, M. (1985). DC-SQUID sytem for high-resolution biomagnetic measurements. Med. & Bio. Eng. & Comput. 23 Suppl., Part 1, 9-10.

Mizutani, Y., and Kuriki, S. (1986). Somatically evoked magnetic fields in the vicinity of the neck. IEEE Trans. Biomed. Eng. BME-33, 510-516.

Nemoto, I., Kazuhito, O., and Toyotama, H. (1985). Magnetometric estimation of intracellular motility of pulmonary macrophages under the influence of inhibitors of ATP synthesis. Med. & Bio. Eng. & Comput. 23 Suppl., Part 1, 54-55.

Okada, Y.C., and Nicholson, C. (1987). Currents underlying the magnetic evoked field in the cerebellum. This volume.

Romani, G.L., Williamson, S.J., and Kaufman, L. (1982a). Biomagnetic instrumentation. Rev. Sci. Instrum. 53, 1815-1845.

Romani, G.L., Williamson, S.J., Kaufman, L., and Brenner, D. (1982b). Characterization of the human auditory cortex by the neuromagnetic method. Exp. Brain Res. 47, 381-393.

Romani, G.L., Leoni, R., and Salustri, C. (1985). Multichannel instrumentation for biomagnetism. In: Hahlbohm, H.D., and Lübbig, H., Eds. SQUID '85: Superconducting Quantum Interference Devices and their Applications, Walter de Gruyter, Berlin, pp. 919-932.

Romani, G.L., and Narici, L. (1986). Principles and clinical validity of the biomagnetic method. Med. Progr. Techn. 11, 123-159.

Roth, B.J., and Wikswo, Jr., J.P. (1985). The magnetic field of a single axon. Biophys. J. 48, 93-109.

Saarinen, M., Katila, T., Nenonen, J., Seppänen, M., and Siltanen, P. (1985). Simulated and measured magnetocardiograms in localized myocardial abnormalities. In: Weinberg, H., Stroink, G., and Katila, T. (1985), Biomagnetism: Applications and Theory, Pergamon Press, New York, pp. 159-163.

Tesche, C.D., Krusin-Elbaum, L., and Knowles, W.D. (1987). Simultaneous measurement of magnetic field and electric potential generated by *in vivo* hippocampal slices. This volume.

Weinberg, H., Stroink, G., and Katila, T., Eds. (1984). Biomagnetism: Applications and Theory Pergamon Press, New York, 499 pp.

Wikswo, Jr., J.P. (1983). Theoretical aspects of the ECG-MCG relationship. In: Williamson, S.J., Romani, G.L., Kaufman, L., and Modena, I., Eds. Biomagnetism: An Interdisciplinary Approach, Plenum Press, New York, pp. 311-326.

Wikswo, Jr., J.P., Barach, J.P., and Freeman, J.A. (1980). Magnetic field of a nerve impulse: First measurements. Science 208, 53-55.

Williamson, S.J., and Kaufman, L. (1981). Biomagnetism. J. Magn. Magn. Mat. 22, 129-202.

Williamson, S.J., and Kaufman, L. (1981). Evoked cortical magnetic fields. In: Erné, S.N., Hahlbohm, H.D., and Lübbig, H., Eds., Biomagnetism, Walter de Gruyter, Berlin, pp. 353-402.

Williamson, S.J., and Kaufman, L. (1987). Analysis of neuromagnetic signals. In: Gevins, A., and Rémond, A., Eds., Handbook of Electroencephalography and Clinical Neurophysiology, Elsevier, Amsterdam, Chap. 14.

Williamson, S.J., Romani, G.L., Kaufman, L. and Modena, I., Eds. (1983). Biomagnetism: An Interdisciplinary Study, Plenum Press, New York, 706 pp.

Williamson, S.J., Pelizzone, M., Okada, Y., Kaufman, L., Crum, D.B., and Marsden, J.R. (1984). Magnetoencephalography with an array of SQUID sensors. In: Collan, H., Berglund, P., and Krusius, M., Eds. Proceedings of the Tenth International Cryogenic Engineering Conference, Butterworth, Guildford, pp. 339-348.

Wood, C.C., Cohen, D., Cuffin, B.N., Yarita, M., and Allison, T. (1985). Electrical sources in human somatosensory cortex: identification by combining magnetic and potential recordings. Science 227, 1051-1053.

Zimmerman, J.E., Beall, J.A., Cromar, M.W., and Ono, R.H. (1987). Operation of a Y-BA-Cu-O rf SQUID at 81 K. Appl. Phys. Lett., in press.

Zimmerman, J.E., and Radebaugh, R. (1978). Operation of a SQUID in a very low-power cryocooler. In: NBS Special Publication 508, National Bureau of Standards, pp. 59-65.

How many memory systems are there really?: Some evidence from the picture fragment completion task

Joan Gay Snodgrass

New York University

Reprint requests should be sent to:

Professor Joan Gay Snodgrass

6 Washington Place, Room 857

New York University

New York, NY 10003

Running head: How many memory systems are there really?

This chapter is to appear in C. Izawa (Ed.), Current issues in cognitive processes: The Tulane Floweree symposium on cognition. Hillsdale, NJ: Erlbaum, in press.

How many memory systems are there really?: Some evidence from the picture fragment completion task

Joan Gay Snodgrass

New York University

Abstract

The present paper reviews Tulving's (1985) most recent proposal for the existence of separate memory modules and their evidentiary bases, reviews the history of the picture fragment completion task, and describes one version of the task which provides a way of measuring all three of Tulving's modules. In the basic task, subjects are trained on a set of fragmented pictures and then, after a brief delay, presented with the repeated training (now old) pictures mixed with an equal number of new pictures. At the end of the test phase, they may be asked to recall both old and new pictures. Skill or procedural learning is defined as improvement between training and new pictures, perceptual learning is defined as improvement between new and old pictures, and episodic learning is defined as recall of the old and new pictures.

Stochastic independence among these three memory components was tested using subjects, items, and subject-items as the units of analysis. The particular pattern of associations and dissociations depended upon both how learning was measured (whether by absolute or relative differences) and by which unit of analysis was used. Alternative explanations of associations and dissociations across tasks are discussed.

MULTIPLE MEMORY SYSTEMS

In 1984, Endel Tulving, in a Distinguished Scientific Contribution Award address presented at the meeting of the American Psychological Association, asked the question: "How many memory systems are there?" (Tulving, 1985). In answer to this question, he proposed a three-tiered system, arranged much like a wedding cake, in what he called a monohierarchical arrangement. The bottom layer, common to both the remaining two, he called "procedural memory"; the second layer, a specialized subsystem of the first, he called "semantic memory"; and the top layer, a specialized subsystem of the second, he called "episodic memory".

The episodic and semantic systems had been introduced more than ten years previously (Tulving, 1972), and all three systems were outlined in his 1983 book Elements of Episodic Memory. What was new in his address was a clarification and modification of the relationships among the three systems, and of acceptable evidential bases for their existence, prompted in part by a series of published discussions in Behavioral and Brain Sciences (Baddeley, 1984; Roediger, 1984; Tulving, 1984). Figure1 shows these three systems, along with descriptions and examples of each level.

Insert Fig 1 about here

Procedural memory is memory for perceptual, motor, and cognitive skills. Procedural memory, according to Tulving, is not declarative in nature but rather is acquired and demonstrated by doing. For example, typing is a procedure or skill in which expertise is demonstrated by doing, not by describing. We do not care if a typist cannot describe where on the keyboard the "r" is located as long as his performance in typing demonstrates that he "knows" where it is. Semantic memory is encyclopedic knowledge about the world which need not be associated with autobiographical time and place of acquisition to be expressed, and is usually expressed by descriptions. Our typist's knowledge that the "r" on the keyboard is the third key from the left on the upper row is an example of semantic memory. Finally, episodic memory is knowledge about the occurrence of events which are tied to the remember's past and thus are autobiographical. The knowledge on the part of our typist that he typed a letter to ABC company yesterday or that he typed the word "representation" three times in the last paragraph are examples of episodic knowledge.

One test of a classification system is that people agree about the proper assignment of a task to a memory system (independent of whether people agree that the ternary classification actually represents different systems). There is general agreement among students of memory about which tasks represent episodic memory. These are all tasks in which subjects must access conscious awareness about an item's prior occurrence, and include recall, cued recall, and recognition tasks. These tasks have been described as explicit memory tasks by Graf and Schachter (1985). There is less agreement, however, about whether a particular task should be assigned to procedural or semantic memory.

Procedural memory tasks include generalized practice effects on a task as exhibited in improved performance (lower error rates or faster times) on novel or nonrepeated items. These include mirror reading of new triads of words (Cohen & Squire, 1980), which is described as "knowing how" by these investigators; reading of transformed text (Kolers, 1976); improvement in picture naming times (Mitchell & Brown, 1987); and picture fragment completion on new items after practice with a prior set (Snodgrass, Smith, Feenan, & Corwin, 1987; Snodgrass, 1987). In addition, Cohen (1984) describes improvement on the puzzle Tower of Hanoi as a procedural learning effect.

Semantic memory tasks include priming effects on perceptual tasks such as tachistoscopic identification (Jacoby & Dallas, 1981) and spelling (Jacoby & Witherspoon, 1982); word and picture fragment completion on repeated items (Warrington & Weiskrantz, 1968); mirror reading of repeated items (Cohen & Squire, 1980), described as "knowing that" by these investigators;

improvement in reading transformed text with repeated sentences (Kolers, 1976); and repetition priming effects in lexical decision (Scarborough, Cortese, & Scarborough, 1977; Scarborough, Gerard, & Cortese, 1979).

Episodic memory tasks include recall or recognition tests which examine knowledge of the learner about when and where he encountered a particular stimulus. Specific examples are recognition memory for primed words which have also been presented for perceptual identification (Jacoby & Dallas, 1981; Tulving, Schachter & Stark, 1982), or recall of when the Tower of Hanoi puzzle was last presented or whether the subject has ever encountered it before (Cohen, 1984).

Although there seems to be some disagreement in the literature about whether a particular task falls in one category or another, this classification scheme seems sufficiently comprehensive so that virtually any task calling on previous experience can be classified into one of the three categories.

We next turn to the question of how we can decide whether these systems are truly independent in the sense that a lower system can operate independently of a higher system (so the procedural system can subserve one task while both the procedural and semantic systems can subserve a second task).

Insert Fig. 2 about here

Figure 2 presents the two patterns of empirical evidence which Tulving proposes can be used as support for these systems. The first pattern, which he terms "functional independence" and which has sometimes been referred to as experimental dissociation, refers to the effects of independent variables on two tasks which are claimed to tap two different memory systems.

Single dissociation refers to the differential effect of one independent variable on each of the two tasks, and double dissociation refers to the differential effect of two independent variables on each of two tasks. There are three levels of dissociation associated with each of the single and double dissociation patterns. Normal dissociation, the most commonly proposed (and observed) is one in which one independent variable has an effect on Task 1 (indicated by a +) and no effect on Task 2 (indicated by a 0). In strong dissociation, the same independent variable has an effect in one direction on Task 1 and an effect in the opposite direction on Task 2. In weak dissociation, the same independent variable has effects on both tasks, in the same direction, but the effect is stronger in one (++) than in the other (+). The corresponding effects for double dissociation are also shown in Figure 2. The experimental variable can be an experimenter-defined manipulation, such as depth-of-processing, modality of presentation, or number of presentations, or a subject-defined variable such as amnesic versus normal performance.

The second type of pattern is called stochastic independence. Here, we look at the relationship between performance on Task 1 and Task 2 with subjects as the unit of analysis (subject correlation), with items as the unit of analysis (item correlation), or with subject-items as the unit of analysis (subject & item correlation).

Again, these correlational patterns can occur in three levels, with normal dissociation showing no correlation, strong dissociation showing a negative correlation, and weak dissociation showing a positive correlation. For subjects, zero correlation between Task 1 and Task 2 means that subjects who do well on Task 1 do neither well nor poorly on Task 2 and vice versa. Thus, performance on the tasks by subjects is normally dissociated. When the correlation is negative, subjects who do well on Task 1 do poorly on Task 2 and vice versa, so subject performance is strongly dissociated. Finally, when the correlation is positive, subjects' performance is not dissociated between the two tasks. Similar patterns can occur for items, and for subject-items. Tulving (1985) has also shown that the same set of data can produce different patterns of correlations across the three units of analysis.

Most experimental reports of dissociation have focussed on functional dissociation and/or stochastic dissociation with subject-items as the units of analysis. For example, Tulving et al. (1982) primed subjects with words, and then measured their performance on the two tasks of word fragment completion and yes/no recognition. Functional dissociation was found between delay interval and performance. Word fragment completion showed no decline in peformance across a one-week interval, while recognition memory declined sharply. Stochastic dissociation at the level of subject-items was also observed. Words which were successfully completed in word fragment completion were not recognized more successfully than uncompleted words, and vice versa.

Mitchell and Brown (1987) studied repetition priming effects in picture naming. They, like Tulving et al. (1982), found single normal functional dissociation for delay interval, in that facilitation in picture naming of repeated pictures showing no decrease across a six-week interval, while recognition memory showed the expected decrease. They also found stochastic independence between naming latencies and recognition memory at the level of subject-items, in that correctly recognized pictures showed no more priming effects than incorrectly recognized pictures.

HISTORY OF THE PICTURE FRAGMENT COMPLETION TASK

The picture fragment completion task is one of several tasks testing what Graf and Schachter (1985; Schacter, 1987) have termed implicit memory. Implicit memory is tested by procedures which do not demand conscious recollection of the prior event whose memory is being tested. Thus, implicit memory for items previously experienced in the picture fragment completion test is exhibited by increased facility in recognizing repeated pictures. That this facilitation can occur without conscious recollection of the prior experience of identifying such pictures is testified by the many reports of patients suffering from organic amnesia who, while denying ever having seen the pictures before, nonetheless show savings on repeated items (Corkin, 1982; Schneider, 1912; Warrington & Weiskrantz, 1968). It is, of course, desirable to ensure that this type of facilitation is not attributable to some generalized practice effect on the task itself; thus, it is desirable to include among the repeated pictures some new (never-before-encountered) items to test for task practice effects.

Training on the prior items can consist of merely presenting the complete pictures and asking subjects to name them, or can consist of actual practice in identifying the items in their fragmented states. Both words and pictures have been used as stimuli in fragment completion tasks. The manner in which stimuli are rendered incomplete varies widely, particularly when words are used as stimuli. For example, words may be fragmented (Warrington & Weiskrantz, 1968), presented with only a stem (Graf, Squire, & Mandler, 1984), presented with several letters missing but with only a single completion (Tulving et al., 1982), presented tachistoscopically (Jacoby & Dallas, 1981), or masked (Marcel, 1983).

There are fewer ways of rendering pictures incomplete. While it is always possible to present pictures tachistoscopically , presenting incomplete pictures is more problematic. Because the components of a picture are less well-defined than those of a word (which have usually been assumed to be the word's component letters), it is necessary to have a theory of picture recognition in order to systematically delete component parts (e.g., Biederman, 1985, 1987). The pictures used by Schneider (1912) to study implicit learning in Korsakoff amnesics were rendered incomplete by subpart deletion, although more recent stimulus sets (Gollin, 1960; Snodgrass et al., 1987; Vokey, Baker, Hayman, & Jacoby, 1986) have usually been constructed by random or quasi-random deletion of picture segments.

There are a number of methods that can be used to produce and test learning effects in picture fragment identification. As we shall see, the particular training procedure used has a profound effect on learning. It is also undoubtedly true that the testing procedure (or more importantly, the similarity between the testing and training procedure) also has a large effect on the magnitude of learning observed. In the next few sections I review the relatively brief literature on the picture fragment completion task, and its relevance to both subject defined and experimentally manipulated variables.

Schneider's Research

According to Parkin's (1982) review of tasks showing learning in organic amnesia, the first investigation of perceptual learning in amnesics was carried out by Schneider (1912) on three Korsakoff syndrome patients. Although Schneider measured memory performance in four tasks, and on a number of other patient groups, including schizophrenics and depressives, we here consider only performance on picture fragment completion for the three Korsakoff patients and a comparison group of 25 young normals. The remaining three tasks were a figure construction task, a picture naming task, and a verbal task in which subjects learned to fill in gaps in prose passages.

The picture fragment completion task used eight objects presented at from 4 to 14 levels of completion. Pictures were rendered incomplete by subpart deletion. Because some pictures were more complex than others, the number of completion levels varied across the set of objects. Subjects were shown the pictures in ascending order of completion (i.e., with the ascending method of limits) and the level at which correct identification was accomplished was recorded, along with their correct and incorrect responses. Although Schneider does not translate these protocols into summary performance measures, I have done so here for the purpose of comparison with subsequent studies. I computed identification thresholds by averaging the levels at which subjects identified each picture.

The 25 young normal subjects (age range: 19-33) were tested only once. The three Korsakoff patients (aged 32, 48, and 51) were given varying numbers of learning trials. The first (J.B.) was tested on six successive days, and then after one month; the second (Ph.F) was tested only twice, the second time after a delay of four months; and the third (S.K.) was tested six times separated by anywhere from four days to two months. Table 1 shows performance on Trial 1 for the 25 normal subjects and three Korsakoff patients by stimulus object, and Figure 3 shows the learning performance for each of the three Korsakoff patients across trials.

Insert Table 1 and Figure 3 about here

It is clear from Table 1 that the two groups of subjects differed markedly in their initial identification performance (an average of one level). On the other hand, it is also clear that the same items that are difficult for the normal subjects are also difficult for the amnesics. Figure 3 shows that all three subjects learned across the repeated trials. Schneider also asked subjects at the end of each learning trial whether they remembered seeing the pictures before. With a few exceptions, subjects' responses were negative.

<u>Gollin's Research</u>

In 1960, Eugene Gollin, searching for a tool for evaluating the course of cognitive development, created a set of fragmented pictures for use in comparative-developmental research. The questions he chose to investigate were, in his words "first, how complete must the representation of a common object be in order that it be recognized; second, to what extent may the completeness of representation required for recognition be reduced as a function of training?" (Gollin, 1960, p. 289).

His stimuli were constructed by quasi-random deletion of fairly large segments of the picture. The deletions were done cumulatively, so that more complete pictures contained all of the segments of less complete pictures. He produced 23 picture series at five levels of completion, although three of the pictures were reserved for practice series. Figure 2 shows selected examples of his pictures at three of the five levels (1, 3, and 5). As can be seen from Figure 4, even the most fragmented pictures are often identifiable, thereby producing unwanted ceiling effects, particularly for adult subjects.

Insert Figure 4 about here

Although the Gollin stimuli have been widely used in neuropsychological research (e.g., Warrington & Weiskrantz, 1968; Corkin, 198~~4~~2), Gollin's empirical results concerning the effects of subject characteristics and training variables have been largely ignored. Accordingly, one purpose of this review is to present these findings and discuss them in the context of the separate memories debate.

Gollin's research, found in five papers published in the early 60's (Gollin, 1960, 1961, 1962, 1965, 1966), was concerned with two issues: the role of developmental (subject) variables, specifically age and IQ, on performance on the task; the role of training (experimental) variables on task performance; and their interaction.

In his 1960 paper, he sought to determine the effect of age and IQ on task performance under a variety of different training manipulations. Identification performance was measured in two ways: by thresholds or by proportion correct. In the threshold method, subjects were shown increasingly more complete pictures by the ascending method of limits until they had correctly identified the picture. A subject's threshold was defined as the ordinal rank of the fragmentation within the series, with "1" representing the most fragmented level and "5", the complete picture. Thus, lower thresholds indicate better performance. In the proportion correct method, subjects were shown only the most fragmented level and scored as correct or incorrect.

Baseline threshold performance in the task improved with age, from 30 months to adult (Experiment 1). Without prior training, children identified only 5% of Level 1 pictures, while adults identified 38% (Experiment 3). These pre-training baseline rates were considered low enough to use identification of Level 1 pictures as a measure of training in subsequent studies.

The first training experiment (Experiment 2) consisted of giving children extensive practice in naming the complete pictures; although the interpolated training procedure had a small effect (about 0.2 level) in reducing thresholds when the pictures were repeated, compared to the control group, simply giving the threshold test twice produced a reduction of approximately 1.1 levels for both control and experimental subjects. Thus, a single training trial with the fragmented pictures had a much larger effect than extensive training with the full pictures.

Subsequent experiments used a single level for training (either the intermediate Level 3 or the full Level 5 stimuli) and tested on Level 1, using the proportion correct measure. Experiments 4a and 4b demonstrated that training with fragmented pictures produced much greater learning than training with the full picture, and also revealed an interesting interaction of the training procedure with IQ and age. Figure 5 shows the effect of training level and the age-IQ variable on identification performance.

Insert Figure 5 about here

For all groups, training with Level 3 pictures produced greater learning than training with Level 5 pictures. However, the older and/or more intelligent children benefitted more from Level 5 training than the younger subjects did. The apparent interaction shown in Figure 5 could, however, be attributable to ceiling effects, as all subjects were close to perfect identification performance under Level 3 training.

In subsequent experiments, Gollin (1961, 1962, 1965, 1966) examined the roles of developmental factors, delay between training and test, and the work expended during training on learning in this task. One fascinating finding is that increases in delay consistently failed to reduce identification performance as long as training was with Level 3 (fragmented) stimuli. A significant decline with delay was only observed either when training was with the full picture or when subjects were young children. Gollin (1962) manipulated the amount of effort expended during training by having subjects learn the entire series of 20 pictures by the method of serial anticipation (i.e., by turning the training procedure into an episodic memory task). Under these "high work" conditions, identification performance actually increased with delay between one minute and one day, as is shown in Figure 6.

Insert Figure 6 about here

Gollin interprets the depressed performance immediately after training to the build-up of reactive inhibition from the intensive learning procedure. Whatever the explanation, here is an example of strong functional dissociation: Implicit memory actually improved with delay while explicit memory, were it to have been measured, would have undoubtedly decreased with delay.

Gollins' finding that training with moderately fragmented pictures produced greater priming than training with complete pictures was the inspiration for Experiment 3, reported here and in

Snodgrass (1987). And Gollin's exploration of the effect of delay interval in implicit memory predates many more recent reports of the immunity or relative immunity of priming facilitation to increases in retention interval.

MEASURES OF IMPLICIT LEARNING

Implicit learning can refer to either procedural learning, in which improvement occurs across presentations of novel material through some kind of generalized task practice effect, or to perceptual learning, in which improvement occurs across presentations of repeated material through item-specific effects (see Schachter (1987) for a historical review of the concept of implicit memory). In order to disambiguate abbreviations, and to be consistent with terminology in Snodgrass (1987), I will hereafter refer to procedural learning as skill learning.

Not all tasks provide the opportunity of measuring both components. For example, when subjects are primed (presented with the complete stimulus) during training, there is no baseline from which to measure skill learning. However, in the typical priming paradigm, subjects are presented with both novel and repeated material during test, so it _is_ possible to measure perceptual learning. And when subjects are repeatedly tested with the same items from session to session, there is no way to measure the skill learning component separately from the perceptual learning component, so the improvement in performance will include both types of learning.

In the following discussion, we will assume that performance measures are available from the training session for training items and from the test session for both old (the repeated training) and new items. This is accomplished by exposing subjects to a subset of stimuli during a training session, and then presenting them with both repeated (old) and nonrepeated (new) items during test. Skill learning is assessed by comparing performance on new items with performance on training items, and perceptual learning is assessed by comparing performance on old items with performance on new items. Such a task was used by Snodgrass et al. (1987) and in Experiments 1 and 4 of Snodgrass (1987).

Given that train, new, and old performance values are available, there are still two considerations to be made in measuring skill and perceptual learning. The first concerns the method used to measure performance during test -- namely, whether the entire range of fragmentations is presented and a threshold determined for each stimulus, or whether a single level of fragmented image is presented and the subject's performance scored as a success (correct identification) or failure (incorrect identification).

Thresholds are usually measured by the ascending method of limits, in which we assume that if a stimulus is identified at a certain level, it will also be identified at all higher levels. The area under the psychometric function constructed under this assumption, weighted by the appropriate level at each point, is the mean threshold, so the threshold measure will sometimes be referred to as the area measure. This contrasts with the single level presentation method, which will sometimes be referred to as the point measure.

Strictly speaking, to compute mean thresholds it is necessary to have an underlying metric at the interval scale level for measuring fragmentation level. In practice, however, the usual metric used is the ordinal number of the fragmentation level (e.g., Gollin, 1960; Snodgrass et al., 1987; Vokey et al., 1986). This makes the mean thresholds noncomparable across experiments. However, because most experiments are concerned with changes in thresholds across an experimental manipulation, the lack of cross-experiment comparability is not usually considered a drawback. As we shall see, translation of means to probabilities of identification is dependent upon assuming that means are computed on ordinal levels; this also make probability of identification subject to the number and difficulty of levels of fragmentation.

The second measurement consideration concerns whether the improvement is measured as an absolute difference between baseline and experimental performance, or whether the improvement is measured as a relative difference. Because subject groups often differ in baseline performance, the choice between absolute and relative measures is an important one.

The factorial combination of how learning is computed from baseline (absolute vs. relative) and test type (area vs. point) produces four possible measures for both skill and perceptual learning. These are described below along with their relevant learning models.

Absolute Measures

Area Measures

Absolute area measures of skill and perceptual learning were used in Snodgrass et al. (1987) and Snodgrass (1987). Skill learning was measured by subtracting train from new thresholds [M(new) - M(train)], and perceptual learning was measured by subtracting new from old thresholds [M(old) - M(new)]. For a more direct comparison with subsequent measures, it will be useful to translate mean threshold values into proportion correct measures.

In the method of ascending limits used to collect the threshold data used here, subjects are given eight opportunities to identify each picture (corresponding to the eight levels of fragmentation available for each picture), in which Level 1 represents the most fragmented image and Level 8, the complete picture. Virtually all subjects identify all pictures by Level 8, so the minimum possible identification score is 1/8 or 0.125. The best possible performance, an identification score of 1.00, occurs when subjects identify the first (most fragmented) level. These considerations, along with the assumption that a stimulus identified at a particular level will be identified at all higher levels, leads to the following transformation for converting mean threshold (M) to proportion identified (P):

$P = [(k + 1) - M]/k$, where k is the number of available levels. For the present data, in which $k = 8$,

$$P = (9-M)/8$$

The relationship between skill and perceptual learning measures expressed as proportions and as means are given below, where SL(a) is skill learning and PL(a) is perceptual learning measured as absolute differences:

$$SL(a) = P(new) - P(train) = [M(train) - M(new)]/k, \text{ and}$$

$PL(a) = P(old) - P(new) = [M(new) - M(old)]/k$

Note that when differences in proportion of identifications are used, the differences in means are divided by the number of levels (k) available for identification. Thus the proportion difference measure is useful for comparing savings across experiments using different numbers of identification levels. All measures used here have been converted to proportions. These can be converted to mean difference measures by multiplying by 8.

Point Measures

When only a single level of fragmented image is presented, the absolute difference measure for skill learning is the difference between the proportion of correctly identified train and new items, and the absolute difference measure for perceptual learning is the difference between the proportion of correctly identified new and old items.

Compared to the area measures, which compute the average differences between the psychometric functions, the point measures are the difference between the psychometric functions at the particular level of fragmentation tested. After all measures are discussed, we present a numerical example to illustrate relationships among them.

The absolute measures assume an additive model of learning in which the effect of training is to add a constant increment of performance regardless of the absolute level of baseline performance, so that the appropriate measure is a difference score. In other words, the model assumes a linear learning function. Snodgrass and Corwin (1987) have tested the additive assumption with data from Snodgrass et al. (1987), with supportive results.

Relative Measures

Despite the success of the absolute measures for the normative data, a great deal of evidence from the learning literature points to a negatively accelerated learning function, in which what is learned on a training trial is a constant proportion of what is left to be learned. This standard learning function can be expressed in terms of probability of identification as follows:

$P(t+1) = P(t) + \beta [(1-P(t)]$, where $P(t+1)$ is the probability of correct identification after training, $P(t)$ is the probability of identification before training, and β is the learning parameter. Solving for β leads to the following equation for learning based on the standard learning equation:

$$\beta = [P(t+1)-P(t)]/[1-P(t)]$$

Applying this measure to skill and perceptual learning leads to the following measures.

<u>Area Measures</u>

Relative skill learning, $SL(r)$, is measured by:

$$SL(r) = [P(new)-P(train)]/[1-P(train)]$$

Relative perceptual learning, $PL(r)$, is measured by:

$$PL(r) = [P(old)-P(new)]/[1-P(new)]$$

The same measures expressed in terms *of means rather* than proportions are:

$$SL(r) = [M(train)-M(new)]/[M(train)-1]$$

and

$$PL(r) = [M(new)-M(old)]/[M(new)-1]$$

These relative measures based on means are similar to the savings measure introduced by Ebbinghaus (1885/1913).

<u>Point Measures</u>

Before presenting the relative point measures, it will be useful to present a numerical example to introduce some notation. This numerical example will be used to illustrate the computation of all four sets of measures.

Insert Table 2 about here

Table 2 presents some simplified data for a stimulus with four levels of fragmentation. The means for train (t), new (n) and old (o) stimuli can be computed by multiplying the proportion of identifications at each level, f, by the level of identification and summing. The mean thresholds for train, new and old stimuli are 3.30, 3.10, and 2.50 respectively. Using the subtraction method employed in our previous research, skill and perceptual learning are 0.20 and 0.60 levels respectively. Converting means to proportions by the transformation $P = (5 - M)/4$ leads to values of P's equal to 0.425, 0.475 and 0.625 for the train, new, and old items. Both absolute and relative measures are shown in Table 2.

If only a single level of fragmented image is presented and we assume that prior presentations of unidentified stimuli have no effect on the ultimate level of identification, then we can compute a point measure for the set of data in Table 2 by using the cumulative proportions, F, of identifications. By assuming that level 2 was presented during training and test, we obtain the point measures of skill and perceptual learning shown at the bottom of Table 2.

It should also be noted that the sum of the differences between the cumulative functions for train and new items, $\Sigma(F(n) - F(t))$, equals the difference in means, $M(t) - M(n)$, and the average difference equals the SL(a) measure, $P(n) - P(t)$. A similar relation holds between the sum and mean difference of the cumulative functions for new and old items and $M(n) - M(o)$ and PL(a).

A test for evaluating absolute and relative models on learning in the picture fragment completion task will be presented in the next section, after the general experimental procedure has been described.

TESTS OF THE SEPARATE MEMORIES HYPOTHESIS

In this section, I will describe some tests of the separate memories hypothesis with data collected from the fragmented pictures task (Snodgrass et al., 1987; Snodgrass, 1987). The task is structured so that in some cases all three types of memory -- skill, perceptual, and episodic -- can be measured.

First I will describe the method for all experiments, and then I will describe specific tests of both functional and stochastic independence among memory tasks.

Method for All Experiments

Apparatus and Stimuli

Stimuli were 150 pictures of objects and animals selected from Snodgrass and Vanderwart (1980), which had been prepared for presentation on the Apple Macintosh computer. The fragmented series were created on the Macintosh computer. The fragmentation algorithm is described more fully in Snodgrass et al. (1987). Briefly, the fragmentation process randomly deleted successive 16x16 pixel blocks from the picture to produce eight levels of fragmented images per stimulus. Before fragmenting, blocks containing black pixels were identified so that only these information-bearing blocks were included as candidates for deletion. The numbers of blocks retained at each level were calculated by an exponential function having the following form:

retained blocks (level) = # total blocks [$\underline{a}$EXP(8-level)],

where $\underline{a}$ was set to .7 and level varied from 8 (complete picture) to 1 (most fragmented). Because the pictures varied in number of total information-bearing blocks, the number of blocks displayed varied across levels. However, the percentage of blocks displayed was constant across items, from 8% at the lowest level to 100% at the complete level. At Level 6, approximately 50% of the total blocks were displayed.

Several fragmentation series were created on-line until an acceptable series was produced. An acceptable series was one in which the overall outline of the picture was preserved at the most fragmented level but critical identifying features of the picture (such as an eye or tail in the picture of an animal) were deleted at low fragmentation levels. This series was then saved for use in the subsequent picture fragment completion task. Two series were created for each picture in Sets 3 and 4 for Experiment 4 of Snodgrass (1987), in which a different fragmentation series was used between training and test.

The resulting fragmented pictures are similar to the Gollin (1960) stimuli in that fairly large areas of the picture are deleted. Saving the actual fragmentations permitted us to present exactly the same fragmentation series in test as in training, a procedure which parallels the Gollin procedure.

Figure 7 presents three examples of the fragmented images at selected levels of completion. The levels are numbered from 1 to 8, where 1 is the most fragmented image and 8 is the complete picture.

Insert Fig. 7 about here

Procedure

Two sets of experimental results will be presented. The first set is from a normative study by Snodgrass et al. (1987), referred to as Experiment 0, and the second set is from a series of four experiments (Experiments 1 through 4) which varied training conditions and is reported in Snodgrass (1987).

The basic task used in Experiment 0 consisted of three phases -- a training phase, a brief delay filled by a distractor test, and a test phase. In the training phase, subjects were given 15 fragmented picture series to identify. Each picture series consisted of eight levels of fragmented images presented for identification by an ascending methods of limits. At the presentation of each fragmented image, subjects attempted to identify it by typing its name on the computer keyboard, or by pressing the return key if they could not identify it. When a subject's name was correct (matched one of the possible names for the picture stored in the computer's lexicon), subjects were informed they were correct and the program proceeded to the next picture in the randomly determined series.

At the end of the training phase, there was a 10-minute delay during which subjects performed a paper-and-pencil cancellation of nine's task. Following the delay, subjects were given the test phase of the experiment. In the test phase, the training (now old) pictures were presented again mixed with an equal number of new pictures. Subjects were not informed that some of the training pictures would be repeated. However, because the number of training items was relatively small and the delay short, virtually all of the subjects were aware that some of the training items were repeated during test.

Subjects exhibited two kinds of learning in this task: skill learning, an improvement on new pictures, and perceptual learning, an improvement on repeated pictures. In the original study, skill learning was measured by subtracting new thresholds from train thresholds, and perceptual learning was measured by subtracting old thresholds from new thresholds. Here we use threshold means transformed to proportions under both the absolute and relative difference methods to measure skill and perceptual learning.

A second purpose of Experiment 0 was to test the equivalence of perceptual thresholds at the three levels of training (train, new, and old) across the five sets of 30 pictures used, and to test the equivalence of the two forms of 15 pictures each which alternately served as train/old and new pictures within each set. Across the 10 set-forms (five sets x two forms), the average skill learning was 0.20 level, and the average perceptual learning was almost 2 levels when measured as threshold differences. However, Experiment 0 also revealed a large effect of form on perceptual identification thresholds. An essential condition of the subtraction method is that the two forms within a set be matched in difficulty because both skill and perceptual learning are measured by comparing performance across forms. Even though the set-forms had been matched for picture variables thought to be important in perceptual identification, based on the Snodgrass and Vanderwart (1980) norms, only two sets (3 and 4) showed no main effect of form and no interaction between form and training. Accordingly, the tests of independence reported here are based on data from subjects receiving Sets 3 and 4 only (a total of 40 subjects from Experiment 0).

Because of the lack of equivalence between forms for Sets 1, 2, and 5, experiments reported in Snodgrass (1987) used only Sets 3 and 4. The experiments were all variations on the basic task in that the same test phase was used as in the basic task, but each differed in the nature of the training phase. In Experiment 1, the training phase was identical to that used in Experiment 0 except that subjects were shown the complete picture after correct identification. Because adding complete picture priming to the basic training procedure produced no increase in perceptual learning, results from Experiment 0 and 1 will be reported together for the subject-based analyses. In Experiment 2, subjects saw no fragmented images during study. Instead, they were presented with the complete picture only and asked to name it. In Experiment 3, subjects during training were exposed to only a single level of fragmented image -- the complete picture (Level 8) or one of two more fragmented picture levels (Levels 3 and 5). Upon presentation of the fragmented or complete image, subjects were required to identify it by typing in its name. If their name for the picture was incorrect, they were given feedback in the form of the picture's name. In Experiment 4, subjects were given the full ascending series to identify during training, but the fragments were changed from training to test to assess the effect of explicit fragment memory. In both Experiments 3 and 4, subjects were asked to recall the names of the pictures after the test session. Thus, for Experiments 3 and 4 (and only for these experiments), performance on an episodic memory task was available.

Comparison of Absolute and Relative Measures of Learning

Before presenting learning measures for all experiments, we first consider what the data from Experiments 0 and 1 can tell us about the plausibility of the two learning models described in the previous section on measurement. The model on which the absolute measures are based assumes that a constant increment is added to performance by the training manipulation, while the model on which the relative measures are based assumes that a constant proportion of what is left to be learned is added. One way of evaluating the two models is to determine the relationship between the probability of identification during training, P(t), and the two measures of perceptual learning, PL(a) and PL(r). If the absolute model is correct, then the correlation between P(t) and PL(a) should be zero, and the correlation between P(t) and PL(r) should be positive. If the relative model is correct, then the correlation between P(t) and PL(a) should be negative and the correlation between P(t) and PL(r) should be zero. In short, the correct model of learning will show dissociation between the baseline measure, P(t), and that model's measure of perceptual learning.

Two sets of correlations were computed. The first set was computed using only the 40 subjects in Experiment 0 receiving Sets 3 and 4 plus all 20 subjects in Experiment 1 ($N = 60$), and the second set was computed using all 100 subjects in Experiment 0 plus the 20 subjects in Experiment 1 ($N = 120$). All subjects in Experiment 0 were used for the second set because train and old performance is measured on the same set-form and thus their comparison is not affected by form differences.

Both sets of correlations showed the same pattern: the correlations between P(t) and PL(a) were -.20 and -.23 for the smaller and larger sets of subjects, ps $>$ .05 and $<$ .025 by one-tailed tests; and the correlations between P(t) and PL(r) were +.27 and +.30 for the same two sets, ps $<$ .05 and $<$ .025 by one-tailed tests. The correlations for the relative model are higher than those for the absolute model, thereby suggesting that the absolute model is correct (although neither model achieves that desired outcome of a zero correlation). However, because of some problems with the absolute measures, we shall use both sets of measures in the subsequent analyses.

Insert Table 3 about here

Results

Table 3 presents the results from the five experiments based on subjects as units of analyses for those memory components available. Experiments 0 and 1 have been combined because full picture priming during training in Experiment 1 had absolutely no effect on performance. For Experiments 0 and 1, both skill and perceptual learning are available for analysis because subjects performed picture fragment completion during training. However, no episodic learning performance was measured.

For Experiment 2, full pictures were presented during training for identification, so no training thresholds were measured nor were measures of recall collected. Thus, only the perceptual learning measures are available for Experiment 2, and so results from this experiment cannot be used to test the separate memories model. I have included Experiment 2 because it provides another test of the relationship between the two measures of perceptual learning for an experimental condition which produces relatively poor performance.

For Experiment 3, one-third of the training pictures were presented as complete (at Level 8), one-third were presented at Level 3, and one-third were presented at Level 5. Because perceptual learning was larger (and identical) for pictures presented at fragmented levels (3 and 5) than for pictures presented as complete, I have divided Experiment 3 into two subexperiments. Experiment 3a presents performance for Levels 3 and 5 priming conditions, and Experiment 3b presents performance for Level 8 priming. Identification performance on Level 8 pictures during training, as might be expected, was virtually perfect; however, performance on Levels 3 and 5 were not, so we use these point measures as estimates of training for Experiment 3a. As can be seen in Table 3, these point measures give comparable levels of training performance to the area measures available for the other experiments. Recall of the pictures was also measured for Experiment 3. Recall proportions are reported separately for train/old pictures (which are consistently higher) and for new pictures. Thus, Experiment 3a provides learning estimates for all three types of memory components, while Experiment 3b provides estimates only for the perceptual and episodic memory components.

Finally, in Experiment 4, train, new, and old thresholds as well as recall proportions are available, and thus Experiment 4 also yields estimates for all three memory components.

Functional Independence

The data shown in Table 3 provide tests of functional independence of the three memory components. Functional independence is supported whenever a particular manipulation has an effect on one memory component, but no (or the opposite) effect on the second. The experimental manipulations which distinguish one experimental condition from another were expected to affect perceptual learning rather than either skill or episodic learning. Accordingly we turn first to comparisons of the perceptual learning components, as measured by either the absolute measure, PL(a), or the relative measure, PL(r).

Perceptual Learning

The experimental conditions produced the following ordering of perceptual learning: Experiment 0/1, in which the full series of fragmented pictures were presented during training, produced the highest amount of perceptual learning; Experiments 2 and 3b, which presented only the complete picture during training, produced the lowest amounts of perceptual learning; and Experiments 3a and 4 produced intermediate amounts of perceptual learning. Experiments 3a and 4 presented fragmented pictures during training, but the training was not optimum either because only one level was shown, as in Experiment 3a, or because although the full sequence was shown, the fragmented images were different from those shown during test, as in Experiment 4. Figure 8 shows the probability of identification for new and old pictures for the five experiments. The differences in slopes among the experiments represent differences among the PL(a) measure.

Insert Figure 8 about here

Analyses of variance on both PL(a) and PL(r) showed that differences among the perceptual learning measures were significant. For PL(a), $\underline{F}(4, 139) = 18.08$, and for PL(r), $\underline{F}(4, 139) = 17.66$, both $\underline{p}$s < .001. Planned comparisons among the means of both measures also revealed exactly the same patterns. Perceptual learning was higher for Experiment 0/1 than for the average of Experiments 3a and 4, which did not differ, and the average of Experiments 3a and 4 were significantly higher than the average of Experiments 3b and 2, which did not differ, $\underline{F}s(1, 139) = 16.81$ and 25.46 respectively for PL(a) and 14.63 and 25.66 for PL(r).

Skill Learning

Differences in skill learning were nonexistent for SL(a) or extremely small for SL(r), which actually was negative in Experiment 3a. The analysis of variance on SL(a) across the three experiments in which it was measured showed absolutely no differences among the experiments, while the SL(r) differences were significant, $\underline{F}(2, 109) = 3.18$, $\underline{p} = .046$. An unplanned comparison between the average of Experiments 0/1 and 4 with Experiment 3a showed the latter had significantly lower skill learning, $\underline{F}(2, 109) = 3.14$, $\underline{p} = .047$. However, the SL(r) measure was somewhat more variable than the SL(a) measure because dividing by a small denominator often magnified effects of outlying negative observations.

Episodic Learning

Only three experimental conditions yielded episodic learning measures. Because Experiments 3a and 3b shared identical new picture recall scores, I only examined differences among recall of the old pictures . There was no overall difference among Experiments 3a, 3b, and 4 in old picture recall, $F(2, 61) = 2.60$, $p = .082$. However, the average of 3a and 4 was significantly different from 3b by a planned comparison, $F(1, 61) = 4.63$, $p = .035$.

In summary, then, the following patterns of learning were observed across the five experimental conditions:

for perceptual learning: Exp 0/1 > Exp 3a = Exp 4 > Exp 3b = Exp 2

for skill learning: Exp 0/1 = Exp 4 ≥ Exp 3a

for episodic learning: Exp 3a = Exp 4 > Exp 3b

Skill learning differed from perceptual learning in showing no superiority of Exp 0/1 over Exp 4 and in showing a somewhat questionable superiority of Exp 4 over Exp 3a. On the other hand, episodic learning showed exactly the ordering as perceptual learning. Accordingly, although skill and perceptual learning were functionally dissociated, episodic and perceptual learning were not: Experimental manipulations which improved perceptual learning (specifically, the presentation of fragmented pictures during training), not only improved subjects' ability to identify those pictures when subsequently presented, but also improved subjects' ability to recall those pictures in the absence of a perceptual cue.

Stochastic Independence

Stochastic independence can be tested in each of three ways -- by subjects, items, and subject-items. Similarly, stochastic independence can be tested among three memory types -- skill, perceptual, and episodic. There are three possible pairings of these memory types -- skill/perceptual; skill/episodic, and perceptual/episodic. All three memory pair types can be tested on subjects and on items. However, only the perceptual/episodic pair type can be tested on subject-items because the same subject must be tested on the same item in both tasks. The logic of measuring skill learning requires that it be expressed on a different set of items from the set used during training when subjects are the unit of analysis, and that it be expressed on a different set of subjects from the set used during training when items are the unit of analysis.

Figure 9 shows the three possible tests of stochastic independence by the three possible pairs of memory systems. Neither skill/perceptual nor skill/episodic comparisons can be made for subject-items so these two cells have been crossed out. The remaining cells show the experiments for which the remaining tests are available. When items served as the unit of analysis, Experiments 0 and 1 were analyzed seperately and Experiment 3a and 3b were combined. Only data from Experiment 4 provided all three comparisons between memory systems for all three types of tests.

Insert Figure 9 about here

Test on Subjects

Tests of stochastic independence by subjects are based on the assumption that, to the extent that these memory systems are separate, we might expect subjects to show little or no correlation between their performance on one memory type (e.g., skill), and a second memory type (e.g., episodic). Indeed, one of the strongest pieces of evidence for the existence of separate systems has been the finding that amnesic subjects show normal or near-normal performance on skill and perceptual memory but show abnormally low performance on episodic memory tasks (e.g., Corkin, 1982; Graf & Schacter, 1985, 1987; Hirst & Volpe, 1982; Jacoby, 1982; Schacter & Graf, 1986; Warrington & Weiskrantz, 1968).

Before examining the pattern of correlations across learning measures, it is important to evaluate the measure-to-measure and task-to-task correlations. If the correlations between two measures of skill or two measures of episodic learning are low, then we would not expect a high correlation across skill and episodic memory.

Insert Table 4 about here

Table 4 presents measure-to-measure and task-to-task correlations across subjects. For all experiments, identification of new and old pictures (the P(n)-P(o) correlation) was positively related. For Experiments 0/1 and 4, but not for Experiment 3a, train-new and train-old correlations were also positive and significant. In Experiment 3a, training performance was measured by a point measure (i.e., by the proportion of Level 3 and 5 pictures identified when presented singly during training), rather than by an area measure.

Although the relationship between the skill and perceptual learning measures rightfully belong to the domain of memory-to-memory relationships, they are presented in Table 4 to illustrate a problem with the absolute measures. The positive correlations among P(t), P(n), and P(o) produced a negative correlation between SL(a) and PL(a) for Experiments 0/1 and 4. Because of the existence of these spurious correlations, and because the relation between the absolute and relative measures of skill and perceptual learning were high (see the SL(a)-SL(r) and PL(a)-PL(r) correlations), I will use the relative measures of skill and perceptual learning in memory-to-memory comparisons.

It is worth noting that the single comparison we have between measures of episodic learning -- the correlation between old and new item recall -- was only significant for Experiment 4, and was actually insignificantly negative for Experiments 3a and 3b. Because of these low correlations between new and old item recall, I will consider each of these measures separately in comparisons of episodic with skill and perceptual learning.

Insert Table 5 about here

Table 5 shows the memory-to-memory correlations for the preferred measures of skill, perceptual, and episodic learning. There was no relationship between skill and perceptual learning, and between perceptual and episodic learning, when the basis of the comparison was by subjects. For the skill-episodic relationship, Experiment 4 showed a high and positive correlation between old item recall and skill learning, but no correlation between new item recall and skill learning. Experiment 3a showed a similar pattern of results, although the positive correlation did not reach significance. In summary, then, with a single exception, the three types of learning were dissociated by subjects. The single exception makes no theoretical sense, as skill learning is an improvement between train and new performance, and the episodic learning measure with which it was correlated was based on old item performance. Accordingly, I attribute this single exception to sampling error.

<u>Tests on Items</u>

Tests of stochastic independence by items followed the same logic as the parallel tests by subjects. They were based on the assumption that, to the extent that these memory systems are separate, we expected an item's learning performance for skill to be unrelated to its learning performance for perceptual or episodic memory.

Item comparisons can be made across a number of different manipulations. Items can be compared across levels of training -- i.e., across the train, new, and old conditions within an experiment. When this is done, comparisons of train with new and new with old performance involves different subjects (because a given subject sees a different subset of new pictures from the subset used for the train/old pictures), whereas comparisons of train with old performance involve the same subjects. Nonetheless, in the normative study (Experiment 0), the correlation between train and new performance was higher ($r = +.82$) than that between train and old performance ($r = +.66$) for the 150 pictures used in the normative study.

Items can also be compared across experiments. Set 3 was common across all experiments, while Set 4 was not used in Experiments 2 and 3. Accordingly performance on the 30 items of Set 3 and the 60 items of Sets 3 and 4 was evaluated separately across the experiments having them in common.

Finally, items can be compared across version of fragmentation. In Experiment 4, two different versions of each fragmentation series were used. Correlations across experiments for the same version at the same level of training were higher than correlations within Experiment 4 across the two different versions. For the 60 items in Sets 3 and 4 common to Experiments 0 and 4, the average of the within-version correlations was +.80, while the average of the between-version correlations was +.40. When only the 30 items in Set 3 common to Experiments 0, 1, and 4 were considered, the within-version correlations averaged +.81, and the between-version correlations, +.38. Because of the large effect of fragmentation version on identification performance, it was decided to treat the two versions of fragmented pictures in Experiment 4 as separate items.

For the item analysis, Experiments 0 and 1 were separated because Experiment 1 only used the 30 items in Set 3. Experiments 3a and 3b were combined because, as only 12 subjects participated in Experiment 3, each item was seen by only two subjects at a given level of fragmentation during training; accordingly, no training measures were computed for items in Experiment 3.

Table 6 presents the memory measures based on items for the five experiments under this new reorganization. Note that data for Experiment 4 were based on 120 items when the perceptual identification measures are determined, because of the separation of the two versions of fragmented images, but were based on only 60 items when the episodic learning measures were computed. This is because subjects saw one version of fragmented image during training and the second during testing, so it was impossible to disentangle the effect of version on episodic learning of old items. Although version could have been separately analyzed for new items, it was not considered worthwhile.

Insert Table 6 about here

Means based on items were virtually identical to those based on subjects except for the relative learning measures. Nonetheless, it was of interest to determine whether the statistical comparisons obtained previously among means based on subjects were replicated when items served as the units of analysis. The item-based perceptual learning measures followed the same ordering of experimental conditions as for subjects: namely, that Experiments 0 and 1 were the best, followed by Experiments 3 and 4, followed by Experiment 2. Analyses of variance on both perceptual learning measures were consistent with this ordering. For both PL(a) and PL(r), the overall differences among experiments were significant, $\underline{F}s(4, 264) = 11.60$ and 21.08 respectively, both $\underline{p}s < .001$. Planned comparisons on both perceptual learning measures showed the following ordering of means:

Exp 0 = Exp 1 > Exp 3 = Exp 4 > Exp 2 (all comparisons at $\underline{p} < .001$).

There were no differences among the experiments in skill learning when measured by SL(a) or SL(r). The significant differences in SL(r) in the subject-based analysis was due to Experiment 3a, which is not represented here.

Episodic learning differences were analyzed in two ways. First, because Experiment 4 used both Sets 3 and 4 (while Experiment 3 used only Set 3), we analyzed the difference between episodic learning for old and new training levels of the items by a mixed design, with experiment as a between-subject variable. With this less sensitive analysis, there was no significant difference between Experiments 3 and 4, $\underline{F}(1, 88) = 3.06$, $\underline{p} = .08$; a large difference between old and new recall performance, $\underline{F}(1, 88) = 66.79$, $\underline{p} < .001$, and no interaction ($\underline{F} < 1$). When only Set 3 from

Experiment 4 was analyzed in a completely within-subjects analysis of variance, Experiment 4 produced better recall performance than Experiment 3, $\underline{F}(1, 29) = 12.57$, $\underline{p} < .01$, there was the same large difference between old and new recall performance, and no interaction. The significant effect of experiment in the second analysis arises in part because in Experiment 4, Set 3 pictures produced better recall performance than Set 4 pictures (.67 vs. .64 overall), so eliminating Set 4 pictures from the analysis enhanced the difference between experiments.

In summary, then, the following patterns of learning were observed across the five experiments, when items were used as units of analysis:

for perceptual learning: Exp 0 = Exp 1 > Exp 3 = Exp 4 > Exp 2

for skill learning: Exp 0 = Exp 1 = Exp 4

for episodic learning: Exp 4 ≥ Exp 3

Again, skill learning was functionally dissociated from perceptual and episodic learning. However, because Experiment 3 was not subdivided into its fragmented and complete picture priming conditions, the similarity of patterns observed for both perceptual and episodic learning by subjects were obscured -- episodic and perceptual learning were no longer parallel, and thus could be said to be partially dissociated.

Insert Table 7 about here

Table 7 presents the correlations between measures and tasks based on items. As for the subject analysis, it was remarkable that the correlations between train and new performance were so high, based as they are on different subjects. Again, it points to the powerful effect of items in determining performance.

There were some differences between the patterns of correlations for the item-based and subject-based analyses. First, the correlations between SL(r) and PL(r) were significantly negative for two of the three experiments, whereas they were not in the subject-based analysis. Second, the correlation between the episodic components for Experiment 4 was stronger for items than for subjects.

Insert Table 8 about here

Table 8 presents correlations across memory tasks. As noted previously, there was a significant negative correlation between the skill and perceptual learning measures for two of the three experiments for which they were available. I take this particular pattern of evidence with some reservations, however, because it is probably an artifact of the measurement procedure. None of the other memory-to-memory correlations was significant, so I conclude that virtually all of the item-based analyses support the separate memories position.

Tests on Subject-Items

Subject-item tests can only be carried out for Experiments 3 and 4 which measured both perceptual and episodic learning, because the same subject's performance must be obtained on the same item for at least two different tasks. Support for the separate memories position is obtained when a difference in a set of subject-items is obtained for one task, but no or the opposite difference is obtained for the second.

The most straightforward comparison is to determine whether identification performance during the course of training is different for recalled than for nonrecalled items, as measured by either of the perceptual learning measures. A relationship between episodic and perceptual learning is demonstrated whenever it can be shown that perceptual learning for recalled items is greater than for nonrecalled items.

In order to facilitate comparisons across Experiments 3 and 4, I combined data from priming Levels 3 and 5 in Experiment 3 (eliminating Level 8), and used only data for Set 3 from Experiment 4. This has the effect of equating perceptual learning measures across the two experiments. The results of a 2 (experiment) by 2 (recall status) by 2 (training level) mixed analysis of variance showed that both the main effect of recall and the recall by training interaction was significant, $F(1, 29) = 4.41$, $p = .04$ and $F(1,29) = 4.10$, $p = .05$. There was no main effect of experiment nor did experiment interact with anything else. Planned comparisons showed that the difference between new picture identification for recalled and nonrecalled items was highly significant, $F(1, 29) = 10.64$, $p = .003$, while the difference between old item identification for recalled and nonrecalled items was not, $F < 1$.

Insert Table 9 and Figure 10 about here

Table 9 and Figure 10 show new and old item performance for recalled and nonrecalled pictures combined across experiments, and Table 9 also shows the perceptual learning measures, based on differences between old and new recalled and nonrecalled pictures by subjects. It is apparent that recalled pictures only differed from nonrecalled pictures for new items, not for old items.

The analysis of the perceptual learning measures showed that perceptual learning was significantly greater for recalled than nonrecalled pictures for PL(a) but not for PL(r), $Fs(1, 30) = 4.75$ and 2.83, $ps = .04$ and $.10$ respectively. It is clear that this difference in perceptual learning occurred because new item identification was worse for recalled pictures, not because old item identification was better. This is an important result, because it shows the importance of measuring both new and old picture identification performance. If only old picture identification had been considered, we would have concluded, along with several other investigators using the method of stochastic independence for subject-items, that perceptual and episodic learning were dissociated. In fact, in Experiment 4, recalled pictures had lower train performance (.54) than nonrecalled pictures (.59), although this difference did not reach significance.

The nature of the relationship between identification performance and recall suggests that when a picture is difficult to identify the first time it is presented (i.e., when it is a training or new item), it is easier to recall. Under this interpretation, the fact that perceptual learning is greater for recalled than for nonrecalled pictures is an artifact of the way in which perceptual learning is measured, and says nothing about the relationship between these two types of memory.

Discussion

Although most of the tests of stochastic independence showed normal dissociaton between pairs of memory components, two showed strong dissociation (a negative correlation) and two showed association (a positive correlation).

Subject-based tests revealed a single positive memory-to-memory correlation, that between skill learning and old item recall for Experiment 4. Subjects who showed most learning between training and new items also showed highest recall for old (but not for new) items. Because that relationship seemed inexplicable, I attributed it to sampling error.

Item-based tests revealed that two of three correlations between skill and perceptual learning were negative, thereby supporting strong dissociation and the separate memories position. There is, however, a measurement problem in comparing skill and perceptual learning. Both learning measures have in common new item identification performance, P(n), which is added to the skill learning measure and subtracted from the perceptual learning measure. When a correlation is computed between a gain score measured from baseline and the baseline measure, the correlation tends to be spuriously negative because it contains the error of measurement of the baseline measurement common to both (Lord & Novick, 1968, p. 73). Although the negative correlations between skill and perceptual learning disappeared for the relative measures SL(r) and PL(r) in the subject-based analyses, they reappeared in the item analysis. Thus these negative correlations between skill and perceptual learning would appear to be due to measurement artifact rather than to strong dissociation between skill and perceptual learning..

Finally, subject-item based tests showed a positive correlation between perceptual and episodic learning. However, this positive association was attributable to a difference in new identification scores, not to a difference in old identification scores. This positive correlation would appear to be spurious as it reflects the influence of effortfulness on recall performance for both old and new items.

Having disposed of all of the evidence <u>against</u> Tulving's proposal that there are separate memory modules, are we now prepared to conclude that these experiments support the separate memories position? The answer to that is, alas, "no." The exercise of showing how all the non-zero correlations should be zero could also be turned around to show how all the zero correlations should really be positive. There are issues of measurement which must be solved before any definitive interpretation of patterns of positive, negative, or zero associations can be interpreted.

Certainly, we have shown that real data can evidence the same sort of variable patterns when evaluated across the three units of analyses that Tulving (1985) demonstrated with artificial data. Patterns of dissociations or associations from subjects to items to subject-items are not predictable. Accordingly, I would argue that dissociation or association effects based on the various units of analysis on which they are computed must be interpreted with caution.

GENERAL DISCUSSION

In this discussion, I will consider various interpretations of dissociations between implicit and explicit memory tasks, and then present a synthesis. Schachter (1987) has reviewed three theoretical views which attempt to account for dissociation betwen implicit and explicit memory. Here I consider each of these in turn in the light of results from the present set of experiments. The three views are activation, processing differences or similarities between encoding and test, and the separate memories position with which this paper began.

First, however, I would like to clarify my usage of certain terms. I use the terms "procedural", "semantic" , and "episodic" to describe different memory tasks, even though by such usage I do not mean to commit myself to the separate memories model. Similarly, I use the terms "priming," "perceptual learning," and "item-specific learning effects" interchangeably to refer to the facilitation observed on repeated items, even though some of these terms are not neutral with respect to the locus of the facilitation effect.

Earlier, I identified priming effects in word and picture fragment completion with the semantic system. Most memory theorists agree that priming is an implicit memory phenomenon, although they disagree about whether it should be attributed to the procedural or semantic systems. Tulving (1985) has suggested that priming effects in word fragment completion might occur in the procedural system, the semantic system, or in yet a fourth system that he whimsically called QM for Question Mark. But because we can clearly distinguish between performance improvements expressed on novel items and performance improvements express on repeated items, this distinction should be reflected in our terms for these two types of learning. Therefore, I identify the first kind of learning with the procedural system and the second kind of learning with the semantic system. Thus procedural or skill learning refers to task practice effects and priming or perceptual learning refers to item practice effects.

Let me examine the basic picture fragment completion task in more detail to clarify my usage of terms and to illustrate the theoretical debate. During the initial study phase, subjects are confronted with a task which is, in my view, relentlessly semantic. They are asked to identify a picture from a few fragments. Such a task clearly relies on semantic memory for its execution. Subjects need to access information about the way objects (or their line drawing representations) look and to find information about the names of these objects. This information has accumulated over the years into the subject's semantic system and at this point in the subject's life, is presumably not associated with a single episode of learning. Thus, this phase of the task is decidedly not an episodic memory task. In addition to accessing information about the appearance and names of specific items, however, subjects must also use some general skills in recognizing objects. Such object recognition skills fall within the procedural memory system.

In the test phase of the experiment, subjects are shown both old and new pictures. When they are confronted with new pictures, their performance shows a small improvement over the training pictures. Somehow, their experience with the training pictures has increased their facility in the task. While the act of identifying new pictures relies on semantic memory in the same way that the act of identifying training pictures did, the <u>increase</u> in facility between training and testing reflects some improvement in the general procedure of object recognition. That is, the <u>improvement</u> is procedural in nature, even though the task on which such improvement is exhibited -- identification of new pictures -- is semantic. I have speculated that this improvement consists of learning to see connections among fragments representing lines and curves of the picture's outline, because subjects appear to need to see some minimal number of fragmented images during training to exhibit such skill learning (Snodgrass, 1987).

When subjects are confronted with old pictures in the test session, they show a much larger improvement compared to the training pictures, the priming or perceptual learning effect. It is the site and nature of this improvement -- the priming or perceptual learning effect -- which is in question.

The activation model of priming (Graf & Mandler, 1984; Mandler, 1980; Morton, 1979) holds that the act of identifying a repeated picture is, like the act of identifying a training or new picture, based upon retrieval of information from semantic memory. However, because this information in semantic memory has recently been accessed, it exhibits increased availability or priming which leads to facilitation in picture identification.

Under the activation model, the priming effect reflects a change in the state of a subject such that an impoverished stimulus which, without prior exposure, would remain a meaningless jumble of lines and curves, comes to more easily evoke its stored semantic memory image. Such evocation occurs without conscious effort on the part of the observer, and indeed may be impeded by effortful recollection. Such a process could be viewed as the lowering of a threshold for the stored visual image, represented either as a set of features or as a prototypical visual image (Snodgrass, 1980, 1984). The fragmented image acts like a key in a lock, producing a perceptible image where otherwise there would be only a meaningless jumble. The "key-in-the-lock" metaphor accords well with phenomenological experience in this task.

The activation explanation places the locus of priming effects squarely in the domain of the semantic system, and thus is consistent with the separate memories view. Schachter (1987) points out two problems with the activation view in accounting for priming effects in normal subjects. The first is the sometimes long-lasting effects of a single priming experience, and the second is the apparent priming effects observed for new associations. The long lasting effects of priming are problematic because, according to Tulving's (1983) characterization, retrieval from semantic memory leaves it unchanged. New associations are presumably formed in episodic memory, and thus priming effects for new associations must occur there if the distinctions between semantic and episodic memory systems are to be maintained.

The processing explanation of priming holds that the similarity of processing operations between study and test determines the amount of learning (Craik, 1983; Jacoby, 1983; Kolers & Roediger, 1984; Roediger & Blaxton, 1987; Roediger & Weldon, 1987). Priming effects are optimal when the test representation most closely matches the study representation. Because priming by its definition relies on presentation of some version of the stimulus for learning to be exhibited, performance in priming tasks has also been described as data-driven, in contrast to episodic tasks such as recall which are conceptually-driven (Roediger & Blaxton, 1987).

The processing similarity explanation strongly resembles the principle of encoding specificity proposed by Tulving (1972, 1973; Tulving & Thomson, 1971) to account for context effects in episodic learning. It differs from encoding specificity in emphasizing processing similarities rather than stimulus context similarities; however, because priming tasks depend upon presenting the same stimulus again, it is hard to distinguish the two positions operationally. The processing similarity explanation differs from both the activation model and the separate memories model in assuming that priming effects reflect the establishment of new episodic representations..

The processing similarity /encoding specificity approach gives a good account of the present data. Perceptual learning was best when the study and test conditions were identical, and was worst when the study and test conditions differed (as in Experiment 2 and 3b when complete pictures were used as primes). Furthermore, it gives a good account of the particular pattern of associations between recall and identification: recall was best when identification was worst -- that is, when subjects on the first presentation of an item needed to do a good deal of work.

To recapitulate, the activation and separate memories position are similar in asserting that priming effects occur in the semantic system (although Tulving has admitted the possibility of yet a fourth system). Their problems are similar in the difficulty of reconciling the extreme longevity of priming with short-lived activation of the semantic system (Corkin (1982) showed that the amnesic H.M. showed priming for the Gollin picture task after 14 years!); and with the fact that priming effects seem to be subject to the same principle of encoding specificity seen in episodic memory. Yet to assert that because priming and episodic memory tasks rely on the same processes and principles, they must occur in the same system means giving up the useful distinction between semantic and episodic memory.

In order to preserve Tulving's distinction between semantic and episodic memory systems, we need to change some of the properties of semantic memory. Specifically, we need to propose that the principle of encoding specificity applies to both semantic and episodic memory. Retrieval from semantic memory under perceptually distinct conditions can leave a context-dependent tag which is easily accessed by redintegrating the initial study contexts. For reasons which are unclear (see Schachter, 1987), such a tag does not seem to contain the self-referential information about time and place of occurrence which would identify it easily as having occurred in the learner's autobiographical past. Thus, retrieval of previously-stored information during a priming task seems often to be unaccompanied by conscious recollection of the episode of learning (and seems never to be accompanied by such recollection in amnesic subjects).

The relative longevity of priming may depend not only on the degree of similarity between the encoding and retrieval episodes, but also the similarity of each of these to other episodes in the subject's day-to-day life. Thus, in the Tulving et al. (1982) experiment, recognition as to whether or not a word is old may suffer over a one-week interval because subjects have experienced similar or identical words during the intervening period. Word fragment completion may not because subjects have not experienced similar word fragments during the intervening period (a crossword puzzle addict, on the other hand, might show a decrement). Similarly, H.M. could show savings in the Gollin pictures because the encoding to retrieval similarity was unique compared to his other experiences during the 14 year interval.

In summary, then, it seems to me that a synthesis of theories might deal with not only the present set of results, but the overwhelming mass of data in the literature on functional and stochastic associations and dissociations. By accepting the important role of encoding-test similarity, by permitting greater context-dependent effects in semantic memory, and by considering not only the encoding-to-test similarity but the encoding/test to other environments similarity, we might be able to reconcile conflicting results and interpretations.

References

Baddeley, A. (1984). Neuropsychological evidence and the semantic/episodic distinction. Behavioral and Brain Sciences, 7, 238-239.

Biederman, I. (1985). Human image understanding: Recent experiments and theory. Computer vision, graphics, and image processing, 32, 29-73.

Beiderman, I. (1987). Recognition-by-components: A theory of human image understanding. Psychological Review, 94, 115-147.

Cohen, N. J. (1984). Preserved learning capacity in amnesia: Evidence for multiple memory systems. In L. R. Squire & N. Butters (Eds.), Neuropsychology of Memory (pp. 83-103). New York: Guilford Press.

Cohen, N. J., & Squire, L. R. (1980). Preserved learning and retention of pattern-analyzing skill in amnesia: Dissociation of knowing how and knowing that. Science, 210, 207-210.

Corkin, S. (1982). Some relationships between global amnesias and the memory impairments in Alzheimer's disease. In S. Corkin, K. L. Davis, J. H. Growdon, E. Usdin, & R. J. Wurtman (Eds.), Alzheimer's Disease: A report of progress (pp. 149-164). New York: Raven Press.

Craik, F. I. M. (1983). On the transfer of information from temporary to permanent memory. Philosophical Transactions of the Royal Society of London, 302, 341-359.

Ebbinghaus, H. (1913). Memory: A contribution to experimental psychology (H. A. Ruger & C. E. Bussenius, Trans.). New York: Teachers College, Columbia University. (Original work published 1885).

Gollin, E. S. (1960). Developmental studies of visual recognition of incomplete objects. Perceptual and Motor Skills, 11, 289-298.

Gollin, E. S. (1961). Further studies of visual recognition of incomplete objects. Perceptual and Motor Skills, 13, 307-314.

Gollin, E. S. (1962). Factors affecting the visual recognition of incomplete objects: A comparative investigation of children and adults. Perceptual and Motor Skills, 15, 583-590.

Gollin, E. S. (1965). Perceptual learning of incomplete pictures. Perceptual and Motor Skills, 21, 439-445.

Gollin, E. S. (1966). Serial learning and perceptual recognition in children: Training, delay, and order effects. Perceptual and Motor Skills, 23, 751-758.

Graf, P., & Mandler, G. (1984). Activation makes words more accessible, but not necessarily more retrievable. Journal of Verbal Learning and Verbal Behavior, 23, 553-568.

Graf, P., & Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic patients. Journal of Experimental Psychology: Learning, Memory, and Cognition, 11, 501-518.

Graf, P., & Schacter, D. L. (1987). Selective effects of interference on implicit and explicit memory for new associations. Journal of Experimental Psychology: Learning, Memory, and Cognition, 13, 45-53.

Graf, P., Squire, L. R., & Mandler, G. (1984). The information that amnesic patients do not forget. Journal of Experimental Psychology: Learning, Memory, and Cognition, 10, 164-178.

Hirst, W., & Volpe, B.T. (1982). Automatic and effortful encoding in amnesia. In L.S. Cermak (Ed.), Human memory and amnesia (pp. 369-386). Hillsdale, NJ: Erlbaum.

Jacoby, L.L. (1982). Knowing and remembering: Some parallels in the behavior of Korsakoff patients and normals. In L. S. Cermak (Ed.), Human memory and amnesia (pp. 97-122). Hillsdale, NJ: Erlbaum.

Jacoby, L. L. (1983). Perceptual enhancement: Persistent effects of an experience. Journal of Experimental Psychology: Learning, Memory, and Cognition, 9, 21-38.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. Journal of Experimental Psychology: General, 110, 306-340.

Jacoby, L. L., & Witherspoon, D. (1982). Remembering without awareness. Canadian Journal of Psychology, 36, 300-324.

Kolers, P. A. (1976). Reading a year later. Journal of Experimental Psychology: Human Learning and Memory, 2, 554-565.

Kolers, P. A., & Roediger, H. L., III (1984). Procedures of mind. Journal of Verbal Learning and Verbal Behavior, 23, 425-449.

Lord, F. M., & Novick, M. R. (1968). Statistical theories of mental test scores. Reading, MA: Addison-Wesley.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. Psychological Review, 87, 252-271.

Marcel, A.J. (1983). Conscious and unconscious perception: Experiments on visual masking and word recognition. Cognitive Psychology, 15, 197-237.

Mitchell, D. B., & Brown, A. S. (1987). Persistent repetition priming in picture naming and its dissociation from recognition memory. Journal of Experimental Psychology: Learning, Memory, and Cognition. In press.

Morton, J. (1979). Facilitation in word recognition: Experiments causing change in the logogen models. In P. A. Kolers, M. E. Wrolstead, & H. Bouma (Eds.), Processing of visible language 1 (pp. 259-268). New York: Plenum.

Parkin, A. J. (1982). Residual learning capability in organic amnesia. Cortex, 18, 417-440.

Roediger, H. L. III (1984). Does current evidence from dissociation experiments favor the episodic/semantic distinction? Behavioral and Brain Sciences, 7, 252-254.

Roediger, H. L. III, & Blaxton, T. A. (1987). Retrieval modes produce dissociations in memory for surface information. In D. S. Gorfein & R. R. Hoffman (Eds.), Memory and cognitive processes:

The Ebbinghaus centennial conference (pp. 349-379). Hillsdale, NJ: Erlbaum.

Roediger, H. L. III, & Weldon, M. S. (1987). Reversing the picture superiority effect. In M. A. McDaniel & M. Pressley (Eds.), Imagery and related mnemonic processes: Theories, individual differences, and applications (pp. 151-174). New York: Springer-Verlag.

Scarborough, D., Cortese, C., & Scarborough, H. (1977). Frequency and repetition effects in lexical memory. Journal of Experimental Psychology: Human Perception and Performance, 3, 1-17.

Scarborough, D. L., Gerard, L., & Cortese, C. (1979). Accessing lexical memory: The transfer of word repetition effects across task and modality. Memory & Cognition, 7, 3-12.

Schacter, D. L. (1987). Implicit memory: History and current status. Journal of Experimental Psychology: Learning, Memory, and Cognition, 13, 501-518.

Schacter, D. L., & Graf, P. (1986). Preserved learning in amnesic patients: Perspectives from research on direct priming. Journal of Clinical and Experimental Neuropsychology, 8, 727-743.

Schneider, K. (1912). Über einige klinisch-pathologische Untersuchungsmethoden und ihre Ergebnisse: Zugleich ein Beitrag zur Psychopathologie der Korsakowschen Psychose. [On certain clinical-pathological methods of research and their results. Together with a contribution to the psychopathology of Korsakoff's psychosis]. Zeitschrift fur Neurologie und Psychiatrie, 8, 553-616.

Snodgrass, J. G. (1980). Toward a model for picture and word processing. In P. A. Kolers, M. E. Wrolstad, & H. Bouma (Eds.), Processing of visible language 2 (pp. 565-584). New York: Plenum.

Snodgrass, J. G. (1984). Concepts and their surface representations. Journal of Verbal Learning and Verbal Behavior, 23, 3-22.

Snodgrass, J. G. (1987). Sources of learning in the picture fragment completion task. Manuscript submitted for publication.

Snodgrass, J. G., & Corwin, J. (1987). Perceptual identification thresholds for 150 fragmented pictures from the Snodgrass and Vanderwart picture set. Unpublished manuscript.

Snodgrass, J. G., Smith, B., Feenan, K., & Corwin, J. (1987). Fragmenting pictures on the Apple Macintosh computer for experimental and clinical applications. Behavior Research Methods, Instruments, & Computers, 19, 270-274.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for naming agreement, familiarity, and visual complexity. Journal of Experimental Psychology: Human Learning and Memory, 6, 174-215.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), Organization of memory (pp. 381-403). New York: Academic Press.

Tulving, E. (1983). Elements of episodic memory. New York: Oxford University press.

Tulving, E. (1984). Precis of elements of episodic memory. Behavioral and Brain Sciences, 7, 223-268.

Tulving, E. (1985). How many memory systems are there? American Psychologist, 4, 385-398.

Tulving, E., Schacter, D.L., & Stark, H. A. (1982). Priming effects in word-fragment completion are independent of recognition memory. Journal of Experimental Psychology: Learning, Memory, and Cognition, 8, 336-342.

Tulving, E., & Thomson, D. M. (1971). Retrieval processes in recognition memory: Effects of associative context. Journal of Experimental Psychology, 87, 116-124.

Warrington, E. K., & Weiskrantz, L. (1968). New method of testing long-term retention with special reference to amnesic patients. Nature, 217, 972-974.

Vokey, J. R., Baker, J. G., Hayman, G., & Jacoby, L. L. (1986). Perceptual identification of visually degraded stimuli. Behavior Research Methods, Instruments, & Computers, 18, 1-9.

Footnote

The author would like to thank Kelly Feenan and Youmin Kim for their help in running subjects and analyzing data. June Corwin and Kelly Feenan provided supportive and illuminating discussions about the interpretations of the results presented here. This research was supported in part by a Research Challenge Fund grant from New York University and a University Research Initiative Program grant from the Air Force Office of Scientific Research.

Figure captions

Figure 1. Tulving's (1985) proposal for multiple memory systems.

Figure 2. Patterns of functional and stochastic dissociation.

Figure 3. Perceptual learning in the three Korsakoff syndrome patients studied by Schneider (1912).

Figure 4. Examples of the Gollin (1960) fragmented images at Level 5 (complete), 3, and 1.

Figure 5. Probability of identification (P) of Level 1 pictures as a function of the level of fragmentation of the training pictures for four groups of subjects (high = high IQ, aver = average IQ). (from Gollin, Experiment 4b, 1960).

Figure 6. Probability of identification (P) of Level 1 pictures for adults and children as a function of delay and the level of fragmentation of the training pictures (from Gollin, 1962).

Figure 7. Examples of fragmented images at Levels 8 (complete), 6, 4, and 2 (from Snodgrass, Smith, Feenan, & Corwin, 1987).

Figure 8. Probability of identification (P) for new and old pictures in the five experiments.

Figure 9. The possible tests of stochastic independence.

Figure 10. Probability of identification (P) for new and old recalled and nonrecalled pictures in Experiments 3 and 4.

Table 1

Comparison of Normal and Korsakoff Identification Performance on Initial Trial (from Schneider, 1912)

	Mean threshold (by levels)	
Picture (max level)	Young normals[a]	Korsakoff patients[b]
pear (4)	1.36	2.33
mouse (8)	2.68	4.33
balance (4)	3.00	2.33
Lily of the valley (5)	3.24	5.00
coffee mill (6)	3.64	4.00
baby carriage (8)	4.36	5.00
lamp (7)	4.56	5.67
watering can (14)	8.00	9.67
Mean	3.86	4.79

[a] N = 25; [b] N = 3

Note: Level is the maximum number of levels of fragmentation available.

Table 2

Examples of the Four Measures of Skill and Perceptual Learning

	Measure							
Level	f(t)	f(n)	f(o)	F(t)	F(n)	F(o)	F(n) - F(t)	F(o) - F(n)
1	.00	.00	.10	.00	.00	.10	.00	.10
2	.20	.25	.35	.20	.25	.45	.05	.20
3	.30	.40	.50	.50	.65	.95	.15	.30
4	.50	.35	.05	1.00	1.00	1.00	.00	.00

M(t) = 3.30; M(n) = 3.10; M(o)=2.50

P(t) = .425; P(n) = .475; P(o) = .625

Σ[F(n) - F(t)] = M(t) - M(n); Σ[F(o) - F(n)] = M(n) - M(o)

Area measures:

SL(a) = P(new)-P(train) = [M(train)-M(new)]/4 = .05

PL(a) = P(old) - P(new) = [M(new)-M(old)]/4 = .15

SL(r) = [P(new)-P(train)]/[1-P(train)]= .087

PL(r) = [P(old)-P(new)]/[1-P(new)]= .286

Point measures (at level 2):

SL(a) = F(new)-F(train) = .05

PL(a) = F(old)-F(new) = .20

SL(r) = [F(new)-F(train)]/[1-F(train)]=.062

PL(r) = [F(old)-F(new)]/[(1-F(new)] = .267

Note: t = train; n = new; o = old. f is the proportion of items identified at each level. F is the cumulated proportion.

Table 3

Memory Measures for the Five Experiments Based on Subjects

		Measure								
Exp	N	P(t)	P(n)	P(o)	SL(a)	SL(r)	PL(a)	PL(r)	EL(o)	EL(n)
0/1	60	.54	.57	.81	.03	.07	.24	.57	N/A	N/A
2	20	N/A	.55	.66	N/A	N/A	.11	.24	N/A	N/A
3a	12	.57	.60	.78	.03	-.04	.18	.46	.69	.45
3b	12	N/A	.60	.70	N/A	N/A	.10	.25	.62	.45
4	40	.52	.55	.72	.03	.06	.18	.40	.73	.53

Note: P(t) = prop train items identified; P(n) = prop new items identified; P(o) = prop old items identified. SL = skill learning; PL = perceptual learning; EL = episodic learning. SL(a) = P(n) - P(t); SL(r) = [P(n) - P(t)]/[1 - P(t)]; PL(a) = P(o) - P(n); PL(r) = [P(o) - P(n)]/[1 - P(n)]. EL(o) = prop old items recalled; EL(n) = prop new items recalled. N/A = not available.

Table 4

Correlations Between Measures and Tasks Based on Subjects

	Experiment				
	0/1	2	3a	3b	4
No. subs	60	20	12	12	40
Measure-to-measure					
SL(a)-SL(r)	.99†	N/A	.88†	N/A	.98†
PL(a)-PL(r)	.83†	.92†	.88†	.96†	.88†
Task-to-task					
P(t)-P(n)	.64†	N/A	.41	N/A	.74†
P(n)-P(o)	.60†	.79†	.68*	.80†	.75†
P(t)-P(o)	.55†	N/A	.06	N/A	.73†
SL(a)-PL(a)	-.36†	N/A	.26	N/A	-.42†
SL(r)-PL(r)	-.01	N/A	.20	N/A	-.25
EL(o)-EL(n)	N/A	N/A	-.22	-.27	.41†

Note: P(t) = prop train items identified; P(n) = prop new items identified; P(o) = prop old items identified. SL = skill learning; PL = perceptual learning; EL = episodic learning. SL(a) = P(n) - P(t); SL(r) = [P(n) - P(t)]/[1 - P(t)]; PL(a) = P(o) - P(n); PL(r) = [P(o) - P(n)]/[1 - P(n)]. EL(o) = prop old items recalled; EL(n) = prop new items recalled. N/A = not available.

*$p < .05$

†$p < .01$

Table 5

Correlations Across Memory Tasks Based on Subjects

	Experiment			
	0/1	3a	3b	4
No. of subjects	60	12	12	40
skill/percep				
SL(r)-PL(r)	-.01	.20	N/A	-.25
skill/episod				
SL(r)-EL(o)	N/A	.45	N/A	.51*
SL(r)-EL(n)	N/A	-.35	N/A	.18
percep/episod				
PL(r)-EL(o)	N/A	-.26	-.26	.11
PL(r)-EL(n)	N/A	.29	-.01	.14

Note: SL = skill learning; PL = perceptual learning; EL = episodic learning. SL(a) = P(n) - P(t); SL(r) = [P(n) - P(t)]/[1 - P(t)]; PL(a) = P(o) - P(n); PL(r) = [P(o) - P(n)]/[1 - P(n)]. EL(o) = prop old items recalled; EL(n) = prop new items recalled. N/A = not available.

* $p < .05$

Table 6

Memory Measures for the Five Experiments (based on items)

		Measure								
Exp	N	P(t)	P(n)	P(o)	SL(a)	SL(r)	PL(a)	PL(r)	EL(o)	EL(n)
0	60	.54	.57	.80	.03	.06	.24	.55	N/A	N/A
1	30	.54	.58	.84	.04	.09	.25	.60	N/A	N/A
2	30	N/A	.55	.66	N/A	N/A	.11	.22	N/A	N/A
3	30	N/A	.60	.75	N/A	N/A	.15	.38	.67	.45
4[a]	60/120	.52	.55	.72	.03	.06	.17	.38	.73	.53

Note: P(t) = prop train items identified; P(n) = prop new items identified; P(o) = prop old items identified. SL = skill learning; PL = perceptual learning; EL = episodic learning. SL(a) = P(n) - P(t); SL(r) = [P(n) - P(t)]/[1 - P(t)]; PL(a) = P(o) - P(n); PL(r) = [P(o) - P(n)]/[1 - P(n)]. EL(o) = prop old items recalled; EL(n) = prop new items recalled. N/A = not available.

[a] The two fragmented series for each pictures are treated as separate items for the skill and perceptual learning measures, but are combined for the episodic measures.

Table 7

Correlations Between Measures and Tasks Based on Items

	Experiment 0	1	2	3	4[a]
No. items	60	30	30	30	60/120
Measure-to-measure					
SL(a)-SL(r)	.98†	.98†	N/A	N/A	.96†
PL(a)-PL(r)	.67†	.68†	.92†	.88†	.89†
Task-to-task					
P(t)-P(n)	.78†	.89†	N/A	N/A	.75†
P(n)-P(o)	.60†	.52†	.80†	.49†	.55†
P(t)-P(o)	.62†	.56†	N/A	N/A	.51†
SL(a)-PL(a)	-.51†	-.45†	N/A	N/A	-.32†
SL(r)-PL(r)	-.12	-.45†	N/A	N/A	-.26*
EL(o)-EL(n)	N/A	N/A	N/A	.22	.60†

*$p < .05$

†$p < .01$

[a] The two fragmented series for each pictures are treated as separate items for the skill and perceptual learning measures, but are combined for the episodic measures.

Table 8

Correlations Across Memory Tasks Based on Items

	Experiment 0	Experiment 1	Experiment 3	Experiment 4[a]
No. of items	60	30	3	60/120
skill/percep				
SL(r)-PL(r)	-.12	-.45†	N/A	-.26*
skill/episod				
SL(r)-EL(o)	N/A	N/A	N/A	-.09
SL(r)-EL(n)	N/A	N/A	N/A	-.04
percep/episod				
PL(r)-EL(o)	N/A	N/A	-.10	.18
PL(r)-EL(n)	N/A	N/A	.16	.09

* $p < .05$

† $p < .01$

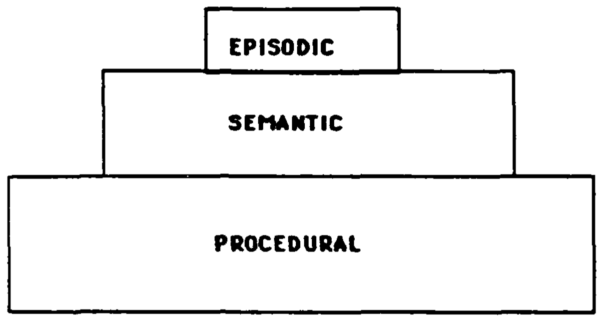
[a] The two fragmented series for each pictures are treated as separate items for the skill and perceptual learning measures, but are combined for the episodic measures.

Table 9

Perceptual Identification and Perceptual Learning for Recalled and Nonrecalled Pictures Combined Across Experiments 3 and 4 (Set 3 Only)

	recalled	nonrecalled
P(n)	.55	.63
P(o)	.76	.76
PL(a)	.21	.13
PL(r)	.47	.33

TULVING'S TRIPARTITE SYSTEM



1. PROCEDURAL "knowing how"

    perceptual, motor, or cognitive skills
    e.g., typing, reading graphs, computer programming, picture recognition
    experimental: nonspecific practice effects on cognitive and motor tasks.

2. SEMANTIC "knowing that" -- implicit memory

    general knowledge shared by a culture
    e.g., knowledge about word meanings, and picture names
    experimental: item specific practice effects on cognitive tasks

3. EPISODIC "remembering that" -- explicit memory

    autobiographical knowledge about self-referenced events
    e.g., knowledge about occurrences of particular words and pictures
    experimental: recognition and recall memory for words and pictures

Figure 1. Tulving's (1985) proposal for multiple memory systems.

Sources of evidence for different memory systems

I. FUNCTIONAL INDEPENDENCE (Experimental)

SINGLE DISSOCIATION

NORMAL

	Task1	Task2
IV 1	+	0

STRONG

	Task1	Task2
IV 1	+	−

WEAK

	Task1	Task2
IV 1	++	+

DOUBLE DISSOCIATION

NORMAL

	Task1	Task2
IV 1	+	0
IV 2	0	+

STRONG

	Task1	Task2
IV 1	+	−
IV 2	−	+

WEAK

	Task1	Task2
IV 1	++	+
IV 2	+	++

II. STOCHASTIC INDEPENDENCE (Correlational)

NORMAL

Task 2	Task 1 Low	Task 1 High
Low	+	+
High	+	+

STRONG

Task 2	Task 1 Low	Task 1 High
Low	0	++
High	++	0

WEAK

Task 2	Task 1 Low	Task 1 High
Low	++	0
High	0	++

(by subjects, items, or subjects & items)

Figure 2. Patterns of functional and stochastic dissociation.

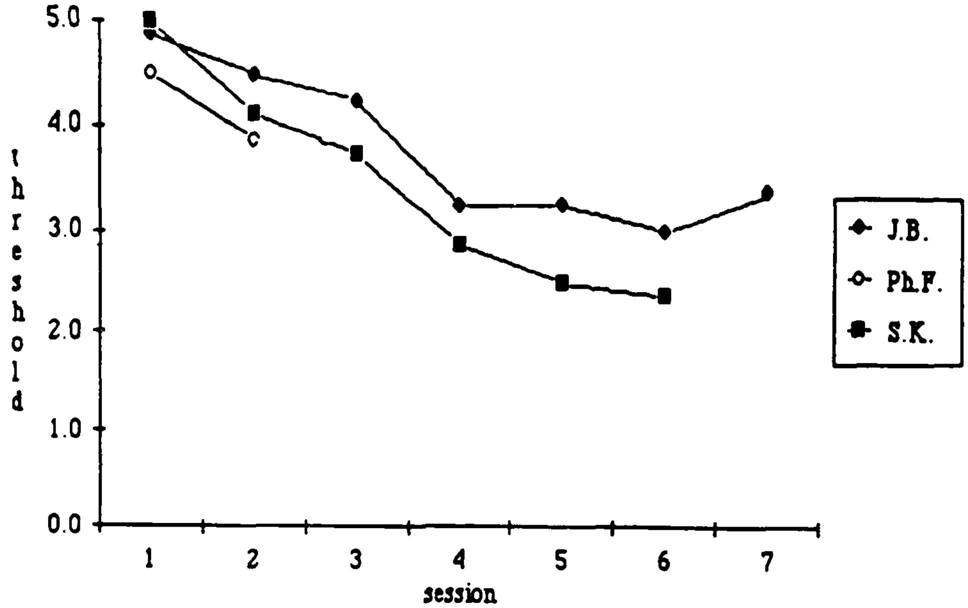


Figure 3. Perceptual learning in the three Korsakoff syndrome patients studied by Schneider (1912).

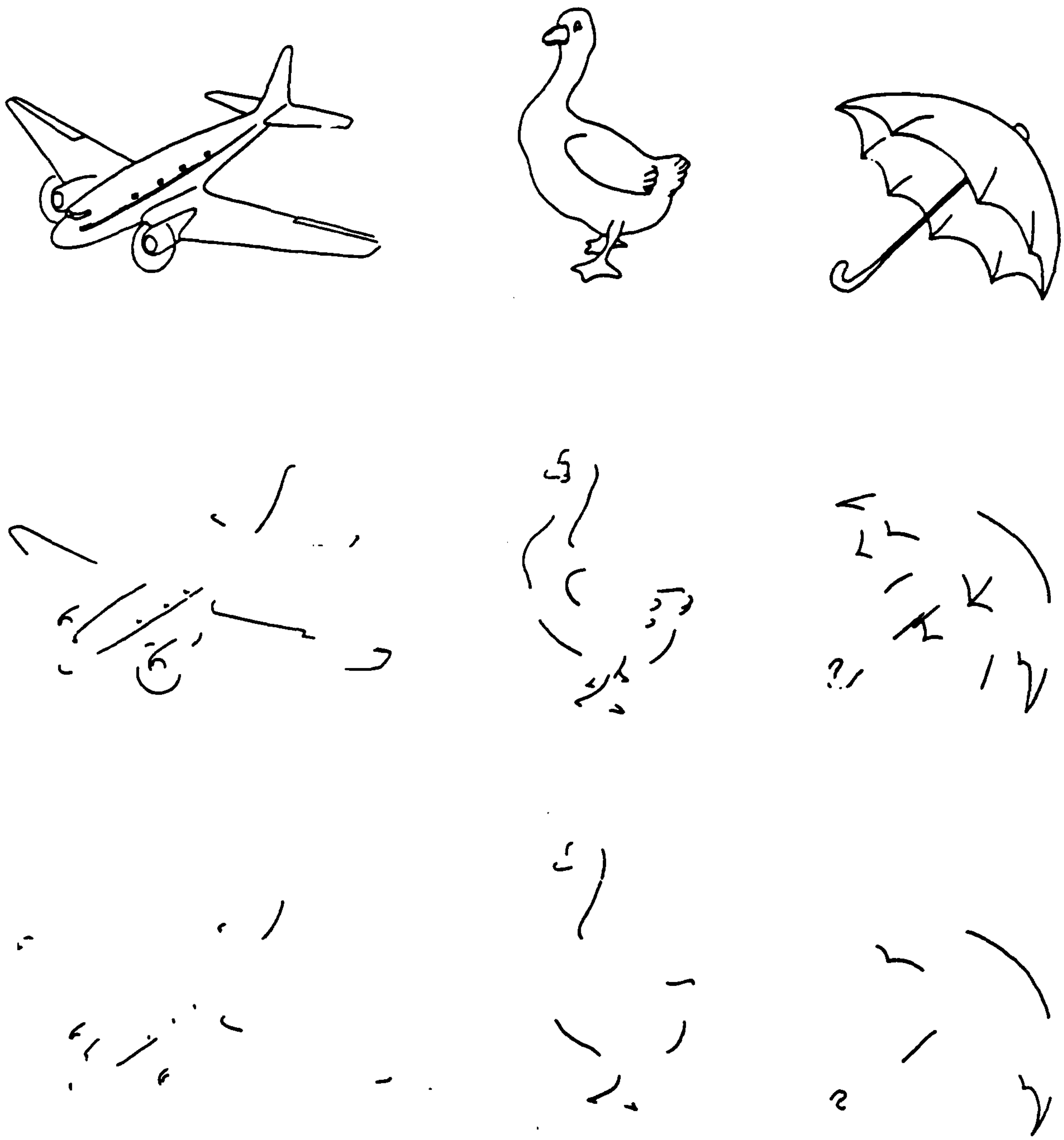


Figure 4. Examples of the Gollin (1960) fragmented images at Level 5 (complete), 3, and 1.

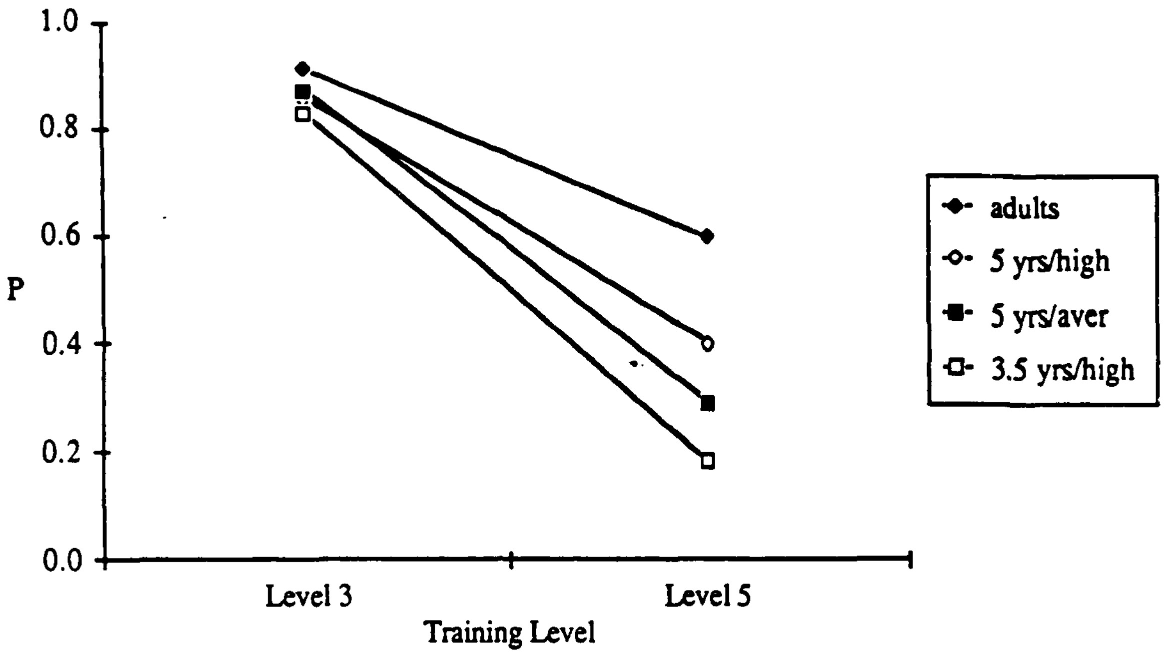


Figure 5. Probability of identification (P) of Level 1 pictures as a function of the level of fragmentation of the training pictures for four groups of subjects (high = high I.Q., aver = average I.Q.). (From Gollin, Experiment 4b, 1960).

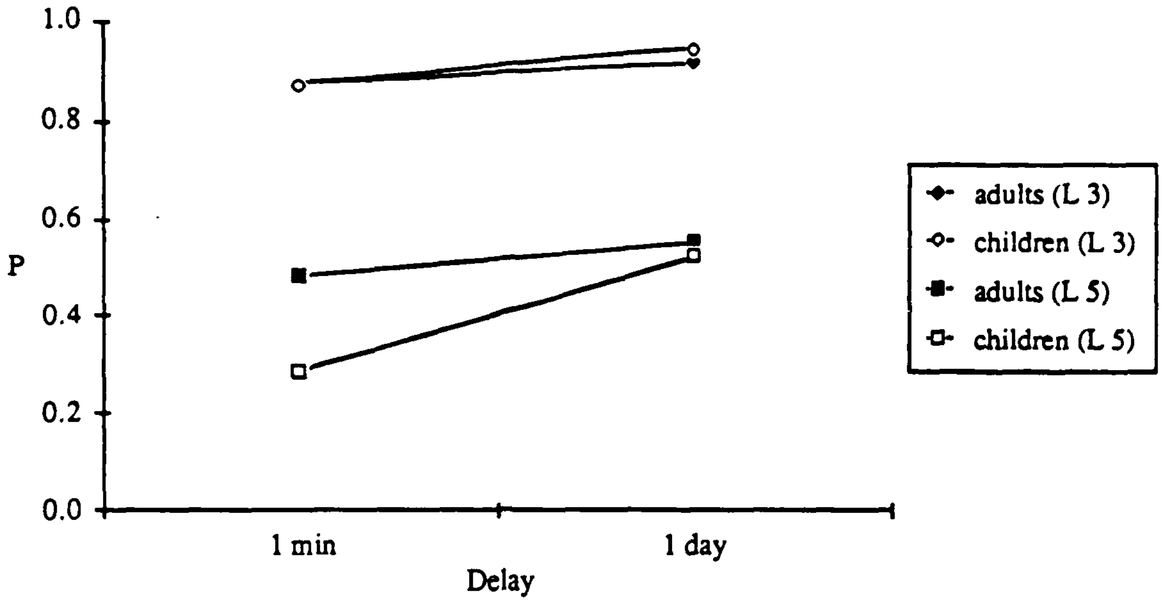


Figure 6. Probability of identification (P) of Level 1 pictures for adults and children as a function of delay and the level of fragmentation of the training pictures (from Gollin, 1962).

Figure 7. Examples of fragmented images at Levels 8 (complete), 6, 4, and 2 (from Snodgrass, Smith, Feenan, & Corwin, 1987).

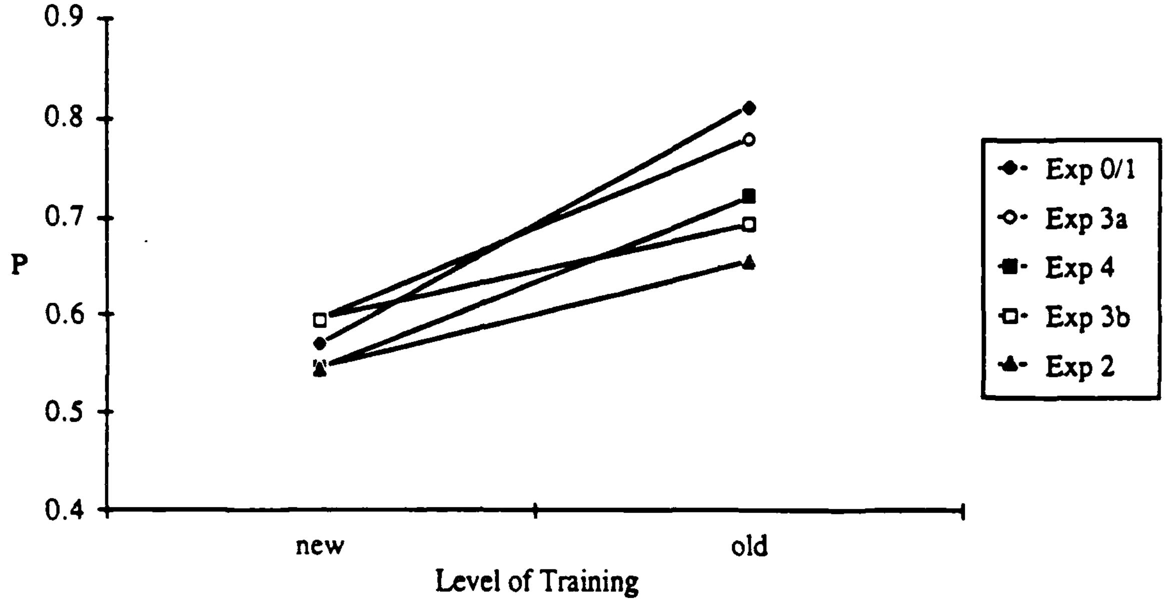


Figure 8. Probability of identification (P) for new and old pictures in the five experiments.

	SUBJECTS	ITEMS	SUBJECT-ITEMS
SKILL / PERCEP	Exps 0/1, 3a, & 4	Exps 0, 1, & 4	
SKILL / EPISOD	Exps 3a & 4	Exp 4	
PERCEP / EPISOD	Exps 3a, 3b, & 4	Exps 3 & 4	Exps 3 & 4

Figure 9. The possible tests of stochastic independence.

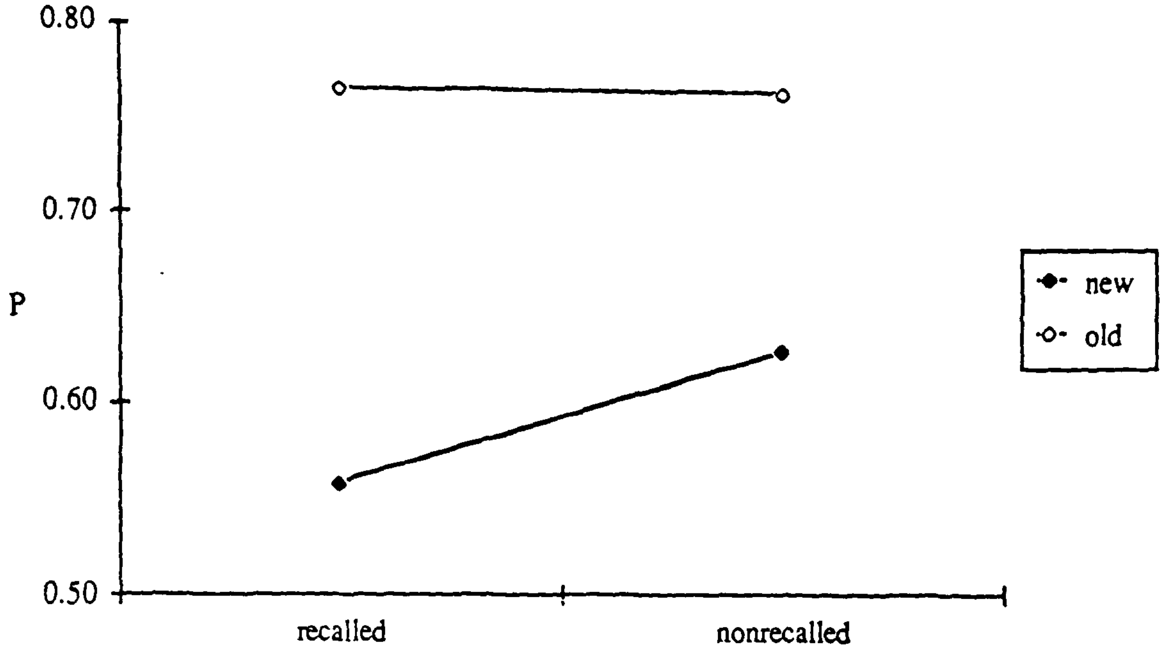


Figure 10. Probability of identification (P) for new and old recalled and nonrecalled pictures in Experiments 3 and 4.

END
DATE
FILMED
DTIC
7-85